UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
PROGRAMA DE MAESTRÍA Y DOCTORADO EN INGENIERÍA
INGENIERÍA ELÉCTRICA - TELECOMUNICACIONES

ESTUDIO Y DESARROLLO DE TÉCNICAS DE PRIVACIDAD GEOGRÁFICA EN REDES INALÁMBRICAS

TESIS
QUE PARA OPTAR POR EL GRADO DE:
DOCTOR EN INGENIERÍA

PRESENTA:
M.I. Oscar Arana Hernández

TUTOR PRINCIPAL
Dr. Javier Gómez Castellanos, Facultad de Ingeniería,UNAM

COMITÉ TUTOR
Dr. Víctor Rangel Licea, Facultad de Ingeniería, UNAM
Dr. Miguel López Guerrero, División de Ciencias Básicas e Ingeniería, UAM-I
Dr. Michael Pascoe Chalke, División de Ciencias Básicas e Ingeniería, UAM-I

MÉXICO, Ciudad de México, enero 2019

JURADO ASIGNADO:

Presidente: Dr. Víctor Rangel Licea
Secretario: Dr. Héctor Benítez Pérez
Vocal: Dr. Javier Gómez Castellanos
1er. Suplente: Dr. Miguel López Guerrero
2do. Suplente: Dr. Michael Pascoe Chalke

Lugar en donde se realizó la tesis: Facultad de Ingernierı́a, UNAM

TUTOR DE TESIS

Dr. Javier Gómez Castellanos

_____

FIRMA

Dedicatoria

A la fuente de mi inspiración, el amor de mi vida, Stephany. Gracias mi amor por esta aventura maravillosa.

A mi madre Gisela, mi padre Raúl y mis hermanos Erin y Rulo, por su apoyo incondicional que siempre me da fuerzas para seguir, los quiero mucho.

Agradecimientos

Agradezco el apoyo brindado a cada una de las personas que laboran en el Posgrado de Ingeniería Eléctrica de la UNAM. Especialmente al Dr. Javier Gómez Castellanos por su trabajo y apoyo incondicional durante todo este proceso. Al Dr. Francisco Garcia Jimenez ya que en todo momento me aconsejó para realizar este trabajo, gracias por la proeza. Asimismo, las aportaciones y recomendaciones que surgieron durante la revisión de este trabajo por parte de: Dr. Miguel López Guerrero, Dr. Víctor Rangel Licea, Dr. Héctor Benítez Pérez y Dr. Michael Pascoe Chalke. Por otro lado, agradezco el apoyo económico brindado por parte del CONACYT y el apoyo recibido dentro del proyecto DGAPA-PAPIIT IN 117017.

Índice general

Resumen **1**

Abstract **3**

1. Introducción **5**
1.1. Definición del problema 6
1.2. Hipótesis 8
1.3. Aproximación al problema 8
1.4. Metas 8
1.4.1. Meta general 8
1.4.2. Metas particulares 8
1.5. Metodología 9
1.5.1. Primera etapa 9
1.5.2. Segunda etapa 9
1.6. Contribución 9
1.7. Descripción del contenido 10

2. Antecedentes **11**
2.1. Estado del arte en las técnicas de privacidad geográfica 11
2.2. Estado del arte en las técnicas de MAC *spoofing* 14
2.3. Resumen del capítulo 15

3. Análisis de TPC **17**
3.1. Resumen del funcionamiento de TPC 17
3.2. Análisis teórico de TPC 18
3.3. Efectos de TPC en los algoritmos de localización 23
3.3.1. Algoritmo de localización CPS 24
3.3.2. Error de estimación del algoritmo CPS 26
3.3.3. Algoritmo de localización WCL 28
3.3.4. Error de estimación del algoritmo WCL 29
3.4. Simulaciones y pruebas en exteriores 30
3.4.1. Simulaciones 30
3.4.2. Pruebas en entornos exteriores 35
3.5. Resumen del capítulo 37

4. MSP **39**
4.1. Análisis de la capa física 39
4.1.1. Modelo del atacante 40
4.1.2. El algoritmo Safe-zone 41

4.2. Análisis de la capa MAC 43
4.2.1. Modelo del atacante 45
4.2.2. El algoritmo VIME 46
4.3. Intercambio de identidad en MSP 48
4.4. Pruebas y experimentos 49
4.4.1. Análisis de seguridad 57
4.4.2. Evaluación de desempeño 58
4.5. Resumen del capítulo 60

5. Conclusiones **61**
5.1. Conclusiones generales 61
5.2. Verificación de la hipótesis 62
5.3. Perspectivas de investigación 63

Glosario **66**

Índice de figuras

3.1. Potencia de transmisión óptima. 20
3.2. TPC_p para $\rho = 0$. 21
3.3. TPC_p para $P_{Tx} = 7\ dBm$ y diferentes valores de σ. 22
3.4. Mapa de valores de TPC_p para $\sigma = 2\,\mathrm{dB}$ y $\rho = 0$. 23
3.5. Mapa de valores de TPC_p para $\sigma = 6\,\mathrm{dB}$ y $\rho = 0$. 23
3.6. Funcionamiento del algoritmo CPS. 26
3.7. Error de estimación del algoritmo CPS en escenario 1-AP. 27
3.8. Funcionamiento del algoritmo CPS cuando nodos móviles implementan TPC. 28
3.9. Funcionamiento del algoritmo WCL. 29
3.10. Funcionamiento del algoritmo WCL cuando nodos móviles implementan TPC. 30
3.11. Mapa de valores de TPC_p para $\gamma = 3$ y $\sigma = 6\,\mathrm{dB}$. 31
3.12. Resultados de simulación para $\gamma = 3$ y $\sigma = 6$. 34
3.13. Error de estimación promedio para diferentes valores de γ y σ. 35
3.14. Prueba en el campus de Ciudad Universitaria. El origen del sistemas de coordenadas se ubica en la esquina superior izquierda (lat $19°19'54.2''$ N, long $99°11'2''$ W). 37

4.1. Funcionamiento del detector KSD. 41
4.2. Candidatos potenciales determinados por Safe-zone para diferentes números de AP. 43
4.3. Relación entre VIME y los componentes de red de un sistema Unix. . . . 47
4.4. CDF de D_{cent}. 51
4.5. Porcentaje de falsos positivos de Safe-zone vs. Pr_{max}. 52
4.6. CDF de los valores P_r obtenidos en el experimento 2. 53
4.7. Candidatos potenciales de Safe-zone cuando el iniciador se encuentra a 15 m con respecto al AP. 54
4.8. Candidatos potenciales de Safe-zone cuando el iniciador se encuentra a 50 m con respecto al AP. 54
4.9. Intercambio de direcciones MAC sin hacer uso de MSP. 55
4.10. Intercambio de direcciones MAC haciendo uso de MSP. 56
4.11. Experimentos de desempeño. 59

Índice de tablas

3.1. Pruebas en escenarios exteriores 36

4.1. Valores de los campos tipo y subtipo en el encabezado IEEE 802.11. . . 46
4.2. Términos utilizados en Safe-zone y VIME. 49
4.3. Valor del umbral $\mathcal{TH}_{attacker}$ vs. el porcentaje de falsos positivos. 50
4.4. Combinación de los detectores de capa física y capa MAC. 56
4.5. MSP vs atacantes. 57
4.6. Evaluación de desempeño. 58

Índice de algoritmos

1. Safe-zone. 44
2. VIME. 48

Resumen

La privacidad geográfica ha sido ampliamente estudiada en el contexto de los servicios basados en la ubicación (LBS). Sin embargo, consideramos que existe otro escenario que representa una mayor amenaza a la privacidad geográfica a los usuarios móviles cuando un conjunto de nodos maliciosos escuchan sus señales para estimar su posición. Esto con el fin de inferir su identidad o hacer uso indebido de esta información sin su conocimiento. Para mitigar este problema, conocido como *location estimation* (LE), se han propuesto algunas estrategias, las cuales se pueden agrupar en: técnicas de ofuscación y técnicas de anonimato.

Por un lado, una de las pocas estrategias de ofuscación propuestas en la literatura con el potencial de proveer privacidad geográfica a los nodos móviles en entornos LE es conocida como control de potencia de transmisión (TPC). En esta técnica, el nodo móvil ajusta su potencia de transmisión de tal manera que únicamente el nodo con el que quiere comunicarse puede escuchar sus transmisiones. Esta acción provoca que los algoritmos de localización sean menos exactos (i.e. la distancia entre la posición real del nodo móvil y la posición estimada por los algoritmos de localización se incrementa notablemente), aumentando el nivel de privacidad geográfica del nodo móvil. Sin embargo, ninguno de los trabajos previos basados en TPC ha evaluado su eficacia considerando factores tales como: límites en la potencia de transmisión del nodo móvil, las condiciones del canal inalámbrico y la densidad de nodos atacantes que escuchan las señales del nodo móvil. Todos estos factores pueden reducir considerablemente la eficacia de esta técnica.

Por otro lado, las técnicas de anonimato utilizan el intercambio frecuente de parámetros relacionados a la identidad del nodo móvil (p. ej. direcciones MAC e IP), para proteger su privacidad geográfica. Este intercambio propicia que los atacantes no puedan relacionar la posición de un nodo móvil con su identidad. Sin embargo, ninguno de los trabajos propuestos en la literatura ha considerado que el intercambio de identidad en una capa puede ser descubierto mediante el análisis de la información en otra capa del modelo de comunicación OSI. Por ejemplo, si consideramos un intercambio de direcciones MAC entre dos usuarios, un atacante, capaz de relacionar la potencia recibida de cada nodo móvil con su posición, podría descubrir el engaño al comparar la similitud entre la potencia recibida anterior y posterior al intercambio. Esto anularía toda protección provista por la técnica de anonimato. Por esta razón, el objetivo de esta tesis es proveer a un usuario móvil de privacidad geográfica aun cuando un atacante tenga acceso a capas uno y dos del modelo de comunicaciones OSI, simultáneamente. Para lograr este objetivo, esta tesis se desarrolla en dos sentidos. Por un lado, se evalúa la viabilidad de usar TPC como técnica de ofuscación para reducir el número de puntos de acceso AP que escuchan al nodo móvil a través de un modelo probabilístico. Por otro lado, se propone MSP, un protocolo que le permite a dos usuarios móviles intercambiar sus direcciones MAC sin ser detectados por

terceros inclusive en los casos en que los atacantes escuchen la capa física y MAC, simultáneamente. En específico, en esta tesis, nos enfocamos en redes basadas en el estándar del IEEE 802.11.

Nuestros resultados muestran que la técnica de ofuscación TPC es insuficiente para garantizar la privacidad geográfica de los nodos móviles en entornos LE. Mientras que nuestras simulaciones y pruebas hechas en entornos exteriores muestran que MSP es capaz de garantizar la privacidad geográfica de los nodos móviles en entornos LE, aun cuando terceros tengan acceso a la información de la capa física y la capa MAC de forma simultánea.

Abstract

Location privacy has been widely studied in the context of location-based services (LBS). However, a far more serious location privacy threat arises when malicious eavesdroppers listen to wireless transmissions from an unsuspecting mobile user in order to pinpoint his location and figure out his identity. This new scenario is known as *location estimation* (LE). While there are several strategies to mitigate the threats posed by LBS scenarios, only a few researchers have dealt with countermeasures for LE scenarios. For instance, Transmission power control (TPC) is one of various obfuscation techniques mentioned in the literature that can provide location privacy to wireless users in LE scenarios. This technique consists of letting the user vary the mobile node's transmission power in a way that only the nearest eavesdropper can overhear the mobile node's signals. This power adjustment has the potential to reduce the number of eavesdroppers overhearing the mobile node's transmissions, thus increasing the location error estimated by nearby eavesdroppers, thereby improving the mobile user's location privacy. However, no previous work has studied its effectiveness as a location privacy technique considering the wireless channel impairments and hardware limitations that can diminish its overall effectiveness. On the other hand, anonymity techniques are the other alternative to provide location privacy to mobile users. In these techniques, a mobile user frequently changes identity parameters, such as MAC and/or IP addresses, or that of other users, making it harder for attackers to associate the location of a transmitting node to its true identity. However, none of previous studies have considered the fact that information from other layers, in particular from the physical layer, might expose a MAC or IP address exchange, thus nullifying the intended location privacy protection.

The goal of this thesis is divided into two steps. In the first step, we analyze the real value of using TPC as a location privacy technique through a probabilistic model that measures the ability of TPC to effectively reduce the number of overhearing eavesdroppers. In the second step, we propose MSP, a MAC swapping protocol that allows two mobile users to discreetly exchange their MAC addresses without malicious eavesdroppers being able to detect it. The results presented in this thesis clearly demonstrate TPC's limited location privacy protection capabilities in most real-life scenarios, disproving previous claims that placed TPC as a solution for the location privacy problem. In contrast, test-bed and simulation experiments demonstrate that MSP is able to ensure location privacy even when attackers eavesdrop on the physical and MAC layers simultaneously.

Capítulo 1

Introducción

Los servicios basados en la ubicación (LBS) se han vuelto cada vez más populares debido al crecimiento y uso exponencial de dispositivos móviles que cuentan con el sistema de posicionamiento global (GPS). Para su correcta operación, estos servicios requieren que el usuario transmita sus coordenadas geográficas con el fin de que este reciba información relacionada con su ubicación como pueden ser la ruta óptima hacia el centro comercial más cercano o el estado del tiempo de alguna ciudad. Sin embargo, el hecho de que el usuario transmita sus coordenadas geográficas puede comprometer su privacidad, ya que terceros o incluso el LBS pueden usar esta información sin su consentimiento y obtener beneficios ilegales. Por ejemplo, terceros pueden predecir la posición futura del usuario y conocer sus hábitos. En la referencia [1], los autores definieron el problema de privacidad geográfica como la capacidad que tiene un usuario de evitar que terceros conozcan su ubicación anterior o actual.

Como contramedida a este problema de privacidad, se han propuesto diversas estrategias las cuales se pueden agrupar en técnicas de ofuscación y técnicas de anonimato. En las técnicas de ofuscación, los usuarios crean posiciones falsas antes de consultar a los LBS. Por otro lado, las técnicas de anonimato desvinculan la posición del usuario de su identidad. Por ejemplo, una de las técnicas más conocidas de anonimato requiere la coordinación entre un grupo de usuarios con el fin de que estos utilicen la misma ubicación para consultar a dichos servicios; esto genera confusión porque el LBS o terceros no pueden distinguir a un usuario del resto. Sin embargo, los autores en la referencia [2] mostraron que ambas técnicas no son efectivas para proteger totalmente la privacidad de los usuarios, en específico cuando se combina información proveniente de varias capas del modelo de comunicación OSI.

Por otro lado, algunos estudios muestran que la posición del nodo móvil puede ser estimada por un conjunto de nodos fijos aun cuando este no transmita su posición explícitamente, usualmente mediante la medición de parámetros inherentes a la comunicación (p. ej. utilizando algoritmos de trilateración [3]). A este problema de privacidad, en esta tesis, lo denominaremos “entornos LE”. Asimismo, al caso en el que terceros obtienen la ubicación de nodos móviles a través de la consulta de LBS lo denominaremos “entornos LBS”.

Es importante resaltar el hecho de que las técnicas de ofuscación y anonimato propuestas para entornos LBS no se pueden usar en entornos LE, ya que en estos últimos la posición del usuario se obtiene de manera indirecta, es decir, sin la participación del usuario. Lo anterior implica que proponer contramedidas que mitiguen el riesgo de ser localizado por terceros (típicamente tres nodos estáticos no colineales) en

entornos LE es completamente distinto a proponer contramedidas en entornos LBS.

Por lo que respecta a entornos LE, también se han propuesto técnicas de ofuscación y anonimato para reducir el riesgo de ser localizado por terceros. La técnica de ofuscación más conocida es el control de potencia de transmisión (TPC). Esta técnica consiste en variar la potencia de transmisión de la interfaz inalámbrica del nodo móvil de tal suerte que solo el punto de acceso (AP) más cercano pueda escuchar la transmisión de dicho nodo. Esta variación tiene el potencial de afectar negativamente la exactitud con la que los algoritmos de localización estiman la posición del nodo móvil. Sin embargo, aun cuando varios trabajos previos han propuesto técnicas de privacidad geográfica basadas en la técnica TPC, ninguno de ellos ha evaluado la eficacia de TPC en entornos más realistas. Por ejemplo, en [4] los autores aluden que la eficacia de TPC está fuertemente relacionada con el número de atacantes, y que dicha efectividad disminuye cuando el número de atacantes aumenta. Sin embargo, los autores no hacen un análisis riguroso de cómo la privacidad geográfica disminuye cuando el número de nodos que detectan al nodo móvil aumenta.

Por otro lado, las técnicas de anonimato para entornos LE suelen cambiar los parámetros que identifican al usuario como son las direcciones MAC o IP [5, 6]. Esto con el fin de que terceros no puedan asociar la verdadera identidad del usuario con su ubicación. Sin embargo, las técnicas de anonimato propuestas para entornos LE tienen la limitante fundamental de que un atacante puede identificar el cambio de parámetros realizado en una capa, mediante el análisis de información en otras capas. Por ejemplo, la información de la capa física (es decir capa uno del modelo de comunicación OSI) podría revelar el intercambio de direcciones MAC. Esto se debe a que un atacante puede relacionar la potencia recibida (RSSI) de un nodo móvil con su posición actual. En consecuencia, un atacante puede identificar cualquier cambio de dirección MAC al comparar la similitud entre el RSSI anterior y posterior al momento del intercambio. Más aún, cuando un nodo móvil cambia su dirección MAC, el sistema operativo de dicho nodo reinicia la interfaz inalámbrica. Esto hace necesario que el nodo se asocie nuevamente con el AP y que el número de secuencia del encabezado IEEE 802.11 se reinicie a cero. Por tanto, las variaciones repentinas de valores de RSSI provenientes de una misma dirección MAC, así como la presencia de paquetes de control relativos a la reasociación con el AP y saltos en los números de secuencias del encabezado IEEE 802.11 generan pistas que indican la utilización de técnicas de anonimato por parte del usuario móvil. Esto anula toda protección a la privacidad geográfica.

Consideramos que los entornos LE, a diferencia de los entornos LBS, representan un mayor riesgo de privacidad geográfica para los usuarios, ya que a diferencia de los entornos LBS, el usuario no sabe que está siendo localizado por un conjunto de nodos maliciosos. Por ello, consideramos que es indispensable proveer a los usuarios de redes móviles con herramientas capaces de garantizar su privacidad geográfica aún en los escenarios más adversos. En específico, en esta tesis, nos enfocamos en redes basadas en el estándar del IEEE 802.11. Sin embargo, los resultados podrían trasladarse a otros estándares (p. ej. IEEE 802.15 o IEEE 802.15.4).

1.1. Definición del problema

Aunque algunos autores han propuesto técnicas de ofuscación y anonimato en entornos LE, consideramos que estas técnicas no garantizan del todo la privacidad geo-

gráfica de los usuarios móviles. Por ejemplo, los trabajos basados en la técnica de ofuscación TPC no consideraron factores tales como: el número de atacantes, el tipo de algoritmo de localización usado por los atacantes y las condiciones del medio inalámbrico. Todo esto genera dudas acerca de la utilidad real de la técnica TPC para proveer privacidad geográfica a usuarios móviles. Por lo anterior, en este trabajo de tesis se propone realizar un análisis riguroso de la viabilidad de usar TPC como contramedida a los algoritmos de localización utilizados por terceros.

Por otro lado, en esta tesis se considera que los usuarios móviles en entornos LE tienen mejores oportunidades de proteger su privacidad geográfica usando técnicas de anonimato. Esto se debe al hecho de que factores como el número de atacantes, el tipo de algoritmo de localización usado por los atacantes y las condiciones del medio no afectan a las técnicas de anonimato tanto como lo hacen con las técnicas de ofuscación. Más aún, debido al hecho de que todos estos factores son ajenos al nodo móvil y que este no tiene ningún control sobre ellos, las técnicas de anonimato proveen una herramienta con la cual el nodo móvil puede tener control de su privacidad.

En particular, esta tesis pretende responder la siguiente pregunta de investigación. ¿Es posible proteger la privacidad geográfica de un nodo móvil mediante el intercambio efectivo de direcciones MAC? La respuesta a esta pregunta está en función de la información proveniente de las primeras dos capas del modelo de comunicación OSI.

Con respecto a la capa física del modelo OSI, en este trabajo de tesis se propone un algoritmo capaz de engañar a los atacantes que detectan las variaciones repentinas de valores de RSSI provenientes de una misma dirección MAC.

Con respecto a la capa MAC del modelo OSI, en esta tesis se propone un algoritmo que intercambia direcciones MAC entre dos usuarios capaz de eliminar la presencia de paquetes de control relativos a la reasociación con el AP y los saltos en los números de secuencias del encabezado IEEE 802.11. Consideramos que este intercambio debe ser únicamente entre dos usuarios, a diferencia de trabajos previos donde la identidad del usuario se protege mediante un grupo de k usuarios que intercambian sus direcciones MAC. La técnica anterior presenta varias desventajas, por un lado encontrar un grupo de $k-1$ usuarios que estén dispuestos a cooperar en el intercambio de direcciones MAC no es tarea fácil, especialmente cuando el valor de k es alto. Por otro lado, debido a que los usuarios deben permanecer desconectados de la red mientras intercambian sus direcciones MAC, el número de nodos móviles dispuestos a cooperar en el intercambio suele ser reducido. También existe en la literatura el enfoque donde el usuario intercambia su dirección MAC por su cuenta. Sin embargo, este enfoque le permite al atacante inferir fácilmente la estrategia del usuario, ya que si un usuario móvil cambia su dirección MAC “A” por la dirección MAC “B”. Desde el punto de vista de los atacantes, la dirección MAC “A” repentinamente desaparece de la red mientras que al mismo tiempo aparece una nueva dirección “B”.

Consideramos que la manera más fiable de proveer privacidad geográfica a los nodos móviles en entornos LE, es a través de combinar una técnica que engañe a los atacantes simultáneamente en ambas capas del modelo OSI. De esta manera se podrá garantizar la privacidad geográfica independientemente de que existan atacantes que escuchen simultáneamente ambas capas.

1.2. Hipótesis

Técnicas de privacidad geográfica basadas en TPC no son efectivas en la práctica debido a factores no estudiados previamente, por el contrario, un solo intercambio de identidad entre dos usuarios que considere la información de la capa fisica y de la capa MAC tiene la capacidad de proveer de privacidad geográfica a usuarios móviles en un entorno LE.

1.3. Aproximación al problema

En entornos LE, los atacantes pueden tener acceso a información del nodo móvil proveniente de varias capas del modelo de comunicación, esto agrava considerablemente el problema de privacidad geográfica debido a que los atacantes puede anular la protección de una capa al considerar la información proveniente de otra capa. En la presente tesis proponemos un protocolo de intercambio de direcciones MAC nombrado MAC *swapping protocol* (MSP). Este protocolo considera la información de la capa física y de la capa MAC simultáneamente con el fin de evitar que los atacantes puedan descubrir el intercambio de direcciones MAC entre dos usuarios. Para lograr este objetivo, MSP utiliza dos algoritmos complementarios. Con respecto a la capa física, se propone un algoritmo denominado *Safe-zone* que se encarga de seleccionar el mejor lugar y el mejor momento para realizar el intercambio de direcciones MAC en capa física. Con respecto a la capa MAC, proponemos un algoritmo denominado *Virtual interface MAC address exchange* (VIME) el cual garantiza que el intercambio de direcciones MAC realizado por dos usuarios pase desapercibido por los atacantes en capa MAC. La combinación de estos algoritmos garantiza la privacidad geográfica de los usuarios móviles en entornos LE.

1.4. Metas

1.4.1. Meta general

Diseñar un conjunto de algoritmos que permita a dos usuarios móviles intercambiar su identidad para garantizar su privacidad geográfica en entornos LE.

1.4.2. Metas particulares

- Realizar un análisis riguroso de la viabilidad de usar TPC como técnica de privacidad geográfica en entornos LE.
- Desarrollar un algoritmo que permita a dos usuarios decidir el mejor lugar y momento para intercambiar su identidad.
- Diseñar un algoritmo que permita el intercambio de direcciones MAC entre dos usuarios que pase desapercibido para los atacantes.

1.5. Metodología

El desarrollo de este proyecto de investigación consta de dos etapas. En la primera etapa se analiza la viabilidad de usar TPC como técnica de ofuscación en entornos LE. En la segunda etapa se propone un protocolo que permite el intercambio de direcciones MAC indetectable en capas uno y dos del modelo OSI.

1.5.1. Primera etapa

Se analiza la viabilidad de TPC mediante un modelo probabilístico, el cual permite a un nodo móvil estimar la potencia de transmisión que maximiza la probabilidad de que su señal inalámbrica alcance únicamente al AP más cercano. Este modelo se evalua bajo diferentes condiciones de densidad de los AP, factores inherentes al canal inalámbrico, diferentes algoritmos de localización utilizados por los atancantes, así como el impacto de utilizar hardware con un intervalo limitado para ajustar su potencia de transmisión.

1.5.2. Segunda etapa

Para proteger la privacidad en capa dos, en este trabajo se desarrolla una técnica de intercambio de direcciones MAC mediante el uso de interfaces virtuales, las cuales permiten al usuario modificar los campos del encabezado IEEE 802.11 antes de entregar los paquetes a la interfaz inalámbrica. Por ejemplo, se puede modificar la dirección MAC de origen y los números de secuencia de los paquetes del usuario, esto permite intercambiar las direcciones MAC sin reiniciar la interfaz inalámbrica y, de esta manera, se elimina su posible detección por parte de terceros.

Por otro lado, para evitar que los atacantes detecten el intercambio mediante información de la capa física, en este proyecto se desarrolla un algoritmo que ayuda al usuario a encontrar un candidato dentro de una zona segura (Safe-zone) para intercambiar sus direcciones MAC. Estas zonas se determinan a través de la potencia de la señal recibida (RSSI) proveniente de los AP. De esta manera, dos usuarios que midan valores similares de RSSI provenientes de los AP pueden intercambiar sus direcciones MAC sin ser detectados por los AP.

1.6. Contribución

- En este trabajo de investigación analizamos la técnica de ofuscación más conocida, TPC. Para ello consideramos factores como: condiciones del canal inalámbrico, algoritmos de localización implementados por los atacantes, así como el impacto de utilizar hardware con un intervalo limitado para ajustar su potencia de transmisión. Como resultado de este análisis se sometió un artículo a la revista *Computer Networks* de la editorial Elsevier.

- Proponemos un algoritmo capaz de identificar el mejor lugar y momento para que dos usuarios intercambien sus direcciones MAC (Safe-zone).

- Proponemos un algoritmo que modifica los paquetes del usuario mediante interfaces virtuales antes de enviarlos a la interfaz inalámbrica (VIME).

- Introducimos el algoritmo MSP, un protocolo capaz de intercambiar direcciones MAC entre dos usuarios sin que sea detectado por terceros, aun cuando estos tengan acceso a las capas física y MAC, simultáneamente. Como resultado de lo anterior, se sometió un artículo que fue aceptado en la revista *Computer Networks* con factor de impacto de 2.5 [7].

1.7. Descripción del contenido

Para alcanzar la meta general plateada previamente, este trabajo de tesis se presenta de la siguiente manera:

- El capítulo 2 presenta una revisión de los trabajos más relevantes respecto a los algoritmos de privacidad geográfica, tanto para entornos LBS como para entornos LE. Asimismo, se incluyen los trabajos más relevantes al robo de identidad (MAC *spoofing*).
- El capítulo 3 presenta un análisis a través de un modelo probabilístico para evaluar la eficacia de TPC. También se incluye el desarrollo de expresiones analíticas que permiten medir el error generado por los algoritmos de localización.
- El capítulo 4 presenta el algoritmo MSP, un protocolo de intercambio de direcciones MAC capaz de pasar desapercibido a atacantes que escuchan las transmisiones inalámbricas en capa física y MAC. Asimismo, se presentan los detalles del diseño e implementación de los algoritmos que conforman a MSP (Safe-zone y VIME). También, se presentan simulaciones y pruebas en entornos exteriores para evaluar su eficacia.
- En el capítulo 5 presentamos las conclusiones generales, la verificación de la hipótesis y las perspectivas de investigación.

Capítulo 2

Antecedentes

En este capítulo se presentan los trabajos más relevantes en cuanto a técnicas de privacidad geográfica en entornos LBS y LE. También se presentan los trabajos más relevantes relacionados con el robo de identidad (MAC *spoofing*).

2.1. Estado del arte en las técnicas de privacidad geográfica

El escenario de privacidad geográfica más estudiado en la literatura considera el caso en el que un usuario móvil consulta información relativa a su ubicación a través de un LBS. El principal problema con este tipo de consulta es que los usuarios móviles deben enviar su ubicación al LBS. Esta actividad puede comprometer la privacidad geográfica del usuario, ya que su ubicación puede ser accesible por terceros o incluso el mismo LBS, quienes pueden hacer uso de esta información sin su consentimiento y obtener beneficios. Como contramedida a este problema, se han propuesto diversas técnicas que se pueden agrupar en técnicas de ofuscación y técnicas de anonimato.

La idea principal en las técnicas de ofuscación es que los usuarios consulten a los servicios basados en la ubicación usando una posición distinta a su posición actual. Esto aumenta la privacidad geográfica del usuario móvil, ya que terceros no conocen la ubicación exacta del nodo móvil [8]. Sin embargo, en la referencia [2], los autores demuestran que se requiere una cantidad considerable de "ruido" en las coordenadas enviadas por el usuario móvil para ocultar su posición actual. En [9], los autores proponen una técnica en la cual el usuario hace referencia a una zona en lugar de su ubicación exacta al momento de consultar a los LBS. Esto le permite al usuario móvil controlar el tamaño y la ubicación de la zona dependiendo del grado de privacidad que requiera. Sin embargo, las técnicas anteriores han demostrado poca eficacia para garantizar la privacidad geográfica y al mismo tiempo obtener un servicio de calidad, ya que ambos objetivos son opuestos.

Por otro lado, la idea principal de las técnicas de anonimato para entornos LBS se basa en cambiar la identidad de los usuarios mediante el cambio frecuente de seudónimos con el fin de ocultar su verdadera identidad [1, 10]. Por ejemplo, una de las técnicas de anonimato más conocida es *spatial cloaking* [11], en la cual un grupo de k usuarios reportan la misma ubicación antes de consultar a los LBS. De esta manera, los LBS no pueden distinguir a un usuario del resto ($k-1$ usuarios). Si bien, estas técnicas confunden a los LBS, su alcance es limitado, ya que terceros pueden conocer

la identidad de los usuarios mediante un análisis de la información recurrente. Por ejemplo, un usuario que constantemente consulta la misma información en los LBS crea indicios que lo vinculan con los diferentes seudónimos que utiliza.

Como se mencionó previamente, las técnicas de ofuscación y anonimato propuestas en los entornos LBS no pueden ser utilizadas de la misma forma en los entornos LE. Esto se debe principalmente a que, en los entornos LE, los atacantes localizan al usuario móvil sin que este tenga conocimiento de ello. Usualmente el proceso de localización se lleva a cabo mediante la medición de parámetros inherentes a la comunicación como son el tiempo de llegada (ToA) [12], diferencia en el tiempo de llegada (TDoA) [13], dirección de llegada (AoA) [14] y potencia recibida (RSSI) [15]. No obstante, el parámetro RSSI es el más utilizado por los algoritmos de localización, ya que su obtención no requiere de hardware especializado.

Para mitigar el problema de los entornos LE, un usuario móvil puede afectar negativamente la exactitud de los algoritmos de localización mediante técnicas de ofuscación. El objetivo de estas técnicas consiste en que los atacantes estimen una posición distinta a la posición real del nodo móvil. Por ejemplo, en [16] los autores demuestran que la exactitud con la que terceros ubican a un nodo móvil está fuertemente relacionada con el número de atacantes (AP); entre más atacantes escuchen la señal inalámbrica del usuario móvil, más exacta será la posición estimada. Como consecuencia de esta relación, en [17] los autores proponen que el usuario varíe la potencia de transmisión (TPC) para reducir el número de AP que escuchan su señal, y así aumentar el error en la posición estimada por terceros. Más aún, los autores en [17] mencionan que hay dos métodos para reducir el número de atacantes en los entornos LE, TPC y el uso de antenas direccionales.

Por un lado, la técnica TPC fue propuesta en un principio como una estrategia para reducir interferencia entre nodos vecinos, así como para ahorrar energía en redes de sensores [18]. Sin embargo, recientemente, algunos trabajos han sugerido su uso como una técnica efectiva para proveer privacidad geográfica a nodos móviles en los entornos LE. TPC, como técnica de privacidad geográfica, fue descrita por primera vez en la referencia [3]. En este trabajo, TPC se describe como una técnica en la cual un nodo móvil constantemente selecciona de manera pseudo-aleatoria la potencia de transmisión con la cual transmitirá su señal inalámbrica. En la referencia [19], los autores cuantifican el efecto de usar TPC para reducir el riesgo de que los atacantes escuchen los paquetes de datos enviados por los usuarios. Sin embargo, aun cuando estos trabajos señalan la viabilidad de TPC como contramedida a los atacantes en entornos LE, su análisis no incluye factores que podrían disminuir la eficacia del método, tales como: el algoritmo de localización utilizado por los atacantes, condiciones del canal inalámbrico y las capacidades limitadas del hardware para ajustar la potencia de transmisión requerida. Recientemente, trabajos como [20] han usado TPC para proveer de privacidad geográfica a los nodos fijos de una red de sensores. En este trabajo, denominado *HyberLoc*, los nodos fijos transmiten paquetes (i.e. *beacons*) encriptados utilizando una potencia de transmisión seleccionada de manera pseudo-aleatoria. Dado que el valor seleccionado de potencia de transmisión es incluido en el *beacon*, solo los nodos que cuentan con la llave podrán estimar la verdadera ubicación de los nodos fijos. Sin embargo, debido a que el intervalo dentro del cual se selecciona la potencia es limitado, los atacantes pueden estimar la ubicación de los nodos fijos después de interceptar un cierto número de *beacons* [21]. Otra estrategia de privacidad geográfica utilizada en redes vehiculares (VANET) fue propuesta en [22]. En este trabajo se considera que la computadora abordo de cada vehículo cons-

tituye un nodo de red. En consecuencia, para proteger la privacidad geográfica de los dispositivos móviles dentro del vehículo, los autores proponen dividir el área de cobertura del nodo de red en células más pequeñas. Cada célula se puede ver como un AP local. De esta manera, los dispositivos móviles que quieran hacer uso de la red deben reducir su potencia de transmisión de tal manera que su señal solo alcance al AP local más cercano. Así, los atacantes ubicados en una célula, no podrán escuchar las señales de otros nodos móviles ubicados en diferentes células. Sin embargo, este trabajo pasa por alto las condiciones del medio inalámbrico, así como la sensibilidad de los receptores a la hora de establecer el tamaño de las células. Estos factores podrían propiciar que la señal inalámbrica sobrepase los límites de la célula o bien que el AP local no pueda decodificar correctamente la señal [23]. Finalmente, cabe resaltar que utilizar una técnica de ofuscación como TPC trae consigo, de forma implícita, el costo relacionado a la posible pérdida de comunicación debido a la utilización de una potencia de transmisión reducida, así como la utilización de la modulación más robusta posible (usualmente BSPK) lo cual reduce considerablemente el *throughput* del usuario. Sin embargo, estos costos deben ser asumidos por el usuario con tal de proteger su privacidad geográfica.

El uso de antenas direccionales es otra técnica propuesta en la literatura con el potencial de reducir el número de atacantes que escuchen las señales inalámbricas de los nodos móviles en los entornos LE. Por ejemplo, en la referencia [24] los autores usan antenas con patrones de radiación directivos para concentrar la señal inalámbrica de un nodo móvil en una dirección en específico, lo que reduce el número de atacantes que pueden recibir dicha señal. Otro ejemplo se encuentra en [25], en el cual los autores construyeron una antena direccional usando una lata de refresco vacía. Esta particular antena generó alteraciones en los algoritmos de localización utilizados en el mismo trabajo. En [26], los autores propusieron una técnica de síntesis del patrón de radiación en antenas direccionales para limitar el número de AP dentro del área de cobertura del nodo móvil. Trabajos más recientes en el campo de Internet de las cosas (IoT) [27, 28], proponen el uso de modelos para medir la probabilidad de que terceros escuchen la señal inalámbrica de los dispositivos de red. Contrario a todos los trabajos anteriores, en este trabajo los autores consideran las condiciones del medio inalámbrico, el número de atacantes, así como el patrón de radiación de las antenas. Sin embargo, solo consideran el uso de antenas direccionales, esto requiere que el nodo utilice hardware especializado y costoso, el cual no se encuentra disponible en la mayoría de los dispositivos comerciales [29].

Por otro lado, las técnicas de anonimato propuestas en entornos LE tienen como objetivo principal que terceros no puedan distinguir a un usuario en particular del resto de los usuarios. Para ello, estas técnicas utilizan el cambio de parámetros que identifican a los usuarios móviles, como son las direcciones MAC e IP. Por ejemplo, en [5] los autores proponen una técnica de anonimato en la que un usuario cambia su dirección MAC cada vez que se asocia a un AP. En [30], los autores proponen una técnica donde cada usuario cambia su dirección MAC tomando como referencia los cambios en la trayectoria y velocidad del nodo móvil. En [6], los autores proponen intercambiar las direcciones MAC de los usuarios a través de un servidor denominado *Dynamic MAC assignment server*, el cual indica a los usuarios qué direcciones usar y cuándo cambiarlas. Sin embargo, una de las principales desventajas a las que se enfrenta un nodo móvil al intercambiar parámetros como su dirección MAC o IP, ya sea por su cuenta o con un grupo, es el hecho de que durante este proceso, el nodo móvil pierde la conexión con el AP [31, 32]. Asimismo, esta interrupción de la

comunicación junto con el proceso de reasociación subsecuente con el AP, le proveen pistas a los atacantes de que los nodos móviles están usando técnicas de anonimato. Para identificar estas pistas los atacantes pueden implementar técnicas denominadas MAC *spoofing*. Estas técnicas son capaces de detectar cuándo un usuario móvil no autorizado le roba la dirección MAC (identidad) a otro usuario móvil legítimo para obtener acceso a la red [33]. Aun cuando las técnicas de MAC *spoofing* no fueron diseñadas para atacar a los nodos móviles en escenarios de privacidad geográfica, estas técnicas pueden ser utilizadas por terceros para detectar cuándo un usuario móvil intenta proteger su privacidad mediante el uso de intercambios de direcciones MAC.

2.2. Estado del arte en las técnicas de MAC *spoofing*

Como se mencionó en el capítulo anterior, si bien el intercambio de direcciones MAC puede confundir a los atacantes, la información proveniente de otra capa del modelo de comunicación OSI puede evidenciar dicho intercambio. Las técnicas de MAC *spoofing* fueron diseñadas para identificar robos de identidad en la capa MAC utilizando información proveniente ya sea de la capa física o la capa MAC. Por ejemplo, el escenario típico asume que dos nodos, uno no autorizado y uno legítimo, transmiten sus señales inalámbricas al mismo tiempo usando la misma dirección MAC. Esto crea variaciones abruptas del parámetro RSSI desde el punto de vista del AP. Lo anterior permite al AP identificar el robo de direcciones MAC por un nodo no autorizado [34]. En [35], por ejemplo, se propuso un detector de MAC *spoofing* el cual crea un perfil para cada usuario mediante un histograma de mediciones RSSI, con esto, el detector es capaz de identificar discrepancias en el comportamiento de un usuario al comparar su histograma previo con el actual. Otro ejemplo lo podemos encontrar en la referencia [36], en la cual los autores proponen un detector basado en el algoritmo *k-means*. Este algoritmo agrupa las mediciones RSSI provenientes de una dirección MAC en k grupos (i.e. clusters) midiendo la similitud entre ellas. De esta manera, si existe un robo de identidad, los clusters se encuentran más separados que cuando las mediciones de RSSI provienen del mismo nodo. Otro enfoque a este problema se presenta en [33], donde los autores utilizan el análisis de Fourier para convertir las mediciones consecutivas de RSSI en el dominio del tiempo al dominio de la frecuencia. Debido al hecho de que durante un robo de identidad las mediciones de RSSI varían de forma drástica, el AP puede identificar el robo mediante la presencia de componentes de alta frecuencia.

Por otro lado, también se han propuesto técnicas de MAC *spoofing* con base en la información de la capa MAC. Estos detectores basan su operación en el análisis del comportamiento del campo *número de secuencia* del encabezado IEEE 802.11 [37]. Este estándar define el número de secuencia como un valor entero que se incrementa en uno por cada paquete que se transmite. Esta operación garantiza la correcta recepción y reensamblado de los paquetes en el receptor. De acuerdo con el estándar, el número de secuencia tiene una longitud de $12\,\mathrm{bits}$, por lo que es posible expresar números enteros entre 0 y 4096. Las técnicas de MAC *spoofing* consideran que cualquier variación del número de secuencia puede indicar un ataque. Por ejemplo, el detector de saltos en el número de secuencia (SNG), propuesto en la referencia [38],

calcula la diferencia entre el número de secuencia del i-ésimo y el i-ésimo - 1 paquetes provenientes de la misma dirección MAC. De esta manera, se considera un ataque de MAC *spoofing* si la diferencia entre estos dos números de secuencia es mayor que un umbral previamente establecido. Otro ejemplo se puede encontrar en la referencia [33], en la cual se describe la técnica de análisis de velocidad de transmisión (SNRA). Esta técnica se basa en el número máximo de paquetes por segundo que una interfaz inalámbrica puede enviar. Por ejemplo, si consideramos una interfaz inalámbrica implementando el estándar IEEE 802.11b, el número máximo de paquetes que dicho estándar puede enviar es 98.214, asumiendo una velocidad de transmisión de 11 Mbit/s. De esta manera, si el número de paquetes provenientes de una dirección MAC excede este número, se considera un ataque de MAC *spoofing*.

2.3. Resumen del capítulo

En este capítulo se presentó una revisión del estado del arte en técnicas de privacidad geográfica en los entornos LBS. Asimismo, se presentaron las técnicas de ofuscación y anonimato más relevantes en los entornos LE. Con respecto a las técnicas de ofuscación basadas en TPC, si bien se presentaron la mayoría de las propuestas existentes, ninguna de ellas ha considerado factores como: condiciones del canal inalámbrico, algoritmos de localización usados por los atacantes, y la potencia limitada de los radios Wi-Fi, entre otros. Consideramos que estos factores pueden limitar la eficacia de usar TPC como técnica de privacidad geográfica. Por otro lado, las técnicas de anonimato más relevantes mostradas en este capítulo no consideran una solución integral que contemple la información de las capas física y MAC durante el intercambio de direcciones MAC. En especial, cuando los atacantes utilizan algoritmos de MAC *spoofing* para evidenciar el intercambio. Por esta razón consideramos necesario el desarrollo de una estrategia de intercambio de identidad que pueda garantizar la privacidad geográfica a los nodos móviles en los entornos LE.

Capítulo 3

Análisis de TPC

En este capítulo, primero se presenta un resumen del funcionamiento general de la técnica de ofuscación TPC. Posteriormente se presenta un análisis de la viabilidad de usar TCP como técnica de privacidad geográfica. Dicho análisis se lleva a cabo mediante un modelo probabilístico que considera factores como las condiciones del canal inalámbrico, diferente densidad de atacantes, diversos algoritmos de localización y, finalmente, el impacto de utilizar hardware con un intervalo limitado para ajustar su potencia de transmisión.

3.1. Resumen del funcionamiento de TPC

La exactitud de los algoritmos de localización está fuertemente ligada con el número de AP que escuchan la señal inalámbrica del nodo móvil. En general, entre más AP escuchen al nodo móvil, mayor será la exactitud con la que estiman su posición. Por ejemplo, en la referencia [16] los autores muestran que hay un incremento significativo en la exactitud del algoritmo de localización cuando dos AP escuchan la señal inalámbrica del nodo móvil en lugar de un solo AP. De la misma manera, la exactitud de la localización aumenta cuando existen tres AP que escuchan la señal inalámbrica comparado con solo dos AP; sin embargo, el incremento en la exactitud no es significativo cuando hay más de tres AP. Como consecuencia de lo anterior, las estrategias de localización suelen usar tres AP para localizar a un nodo móvil. Por el contrario, TPC tiene como objetivo minimizar el número de AP que escuchan la señal inalámbrica de los nodos móviles como mecanismo para reducir la exactitud de los algoritmos de localización usados por los atacantes. En específico, TPC logra este objetivo mediante ajustes a la potencia de transmisión.

En general, podemos considerar que existen dos estrategias para controlar la potencia de transmisión. En el primer enfoque, el nodo móvil selecciona la potencia de transmisión de manera pseudo-aleatoria [17]. Sin embargo, en esta técnica el nodo móvil no tiene ningún control sobre el número de AP que escuchan su señal inalámbrica. En el segundo enfoque, el nodo móvil transmite con la potencia mínima necesaria para alcanzar solo al AP más cercano. De esta forma el nodo móvil tiene control del número de AP que lo están detectando [16]. Por ejemplo, en [3] los autores proponen una estrategia para controlar la potencia de transmisión al medir la diferencia entre la potencia recibida por el AP más cercano y el siguiente en orden decreciente. Si esta diferencia es mayor a un umbral, el nodo móvil transmite con la potencia mínima, caso contrario, transmite con la potencia máxima. Sin embargo, las variaciones

abruptas de la potencia recibida pueden alertar a los AP de que el nodo móvil está usando estrategias de ofuscación. Aun cuando este comportamiento no forma parte del comportamiento estándar de un AP, en este trabajo, lo consideramos como parte del modelo de atacante. Para evitar cualquier indicio que pueda evidenciar la estrategia de ofuscación, TPC debe considerar ajustes graduales de tal suerte que la potencia recibida no presente cambios abruptos entre un paquete y el siguiente. A partir de este punto, al escenario en el cual el nodo móvil usa la potencia mínima necesaria para alcanzar solo al AP más cercano se denominará escenario 1-AP.

En este capítulo se presenta un análisis de la capacidad de TPC para lograr que los algoritmos de localización generen el mayor error posible. En otras palabras, cuantificar la capacidad de TPC para lograr que el algoritmo de localización se encuentre en el escenario 1-AP la mayor parte del tiempo. Sin embargo, para realizar los ajustes de potencia, el nodo móvil debe considerar condiciones del canal inalámbrico como son: factores de desvanecimiento de la señal, pérdida de energía debido a la trayectoria, la sensibilidad de los receptores, así como el número de AP y el algoritmo de localización utilizado por los atacantes. Además, dicho nodo móvil debe ajustar su potencia sin la ayuda de algún otro nodo móvil, ya que esta información podría darle indicios a los atacantes del uso de una estrategia de ofuscación por parte del nodo móvil.

3.2. Análisis teórico de TPC

El análisis de TPC en este trabajo se divide en dos pasos. En el primer paso, un nodo móvil escucha el medio inalámbrico para estimar el número de AP en su vecindad y a su vez estimar qué tan alejados se encuentran con respecto a él. Como resultado de este paso, el nodo móvil obtiene un arreglo $D = \{d_1, d_2, \ldots, d_n\}$ que contiene n distancias estimadas, en el tiempo t. Posteriormente, este arreglo es ordenado de forma ascendente, de tal manera que $\hat{d_1} < \hat{d_2} < \ldots < \hat{d_n}$. Por lo que, $\hat{d_1}$ representa la distancia estimada al AP más cercano. En el segundo paso, el nodo móvil estima el valor de potencia de transmisión que debe utilizar con el fin de que solo el AP más cercano (i.e. el AP ubicado a la distancia $\hat{d_1}$) pueda decodificar correctamente sus paquetes transmitidos. Este problema puede ser formulado de la siguiente manera:

$$TPC_p = p(P_{Rx}(\hat{d_1}) \geq P_{min}, P_{Rx}(\hat{d_2}) < P_{min}) \ldots, P_{Rx}(\hat{d_n}) < P_{min}) \tag{3.1}$$

En esta ecuación, P_{min} es la potencia de recepción mínima para que una señal inalámbrica pueda ser decodificada[1] (i.e. sensibilidad del receptor), $P_{Rx}(\hat{d_1})$ es el valor de la potencia recibida a la distancia $\hat{d_1}$, y $P_{Rx}(\hat{d_1}) \geq P_{min}$ indica el evento en el cual la potencia recibida a la distancia $\hat{d_1}$ se encuentra por encima de P_{min}. En otras palabras, esta ecuación modela la probabilidad de que las señales enviadas por el nodo móvil alcancen al AP ubicado a la distancia $\hat{d_1}$, y al mismo tiempo los AP ubicados a las distancias $\hat{d_i},\ \forall i \geq 2$ no puedan decodificar correctamente la señal inalámbrica del nodo móvil. En este modelo, podemos observar que si la probabilidad del evento $P_{Rx}(\hat{d_2}) < P_{min}$ es cercana a 1, los eventos $P_{Rx}(\hat{d_i}) < P_{min}, \forall i > 2$ también tendrán un valor cercano a 1. Esto se debe al hecho de que la potencia de la señal inalámbrica decrece a medida que la distancia aumenta. Por el contrario, si la probabilidad del evento $P_{Rx}(\hat{d_2}) < P_{min}$ es cercana a 0, indica que más de un AP pueden escuchar la

[1] En este trabajo se asume que todos los nodos inalámbricos tienen la misma sensibilidad.

señal del nodo móvil, por tanto bastará con conocer si la señal puede o no ser decodificada por el AP ubicado en la distancia $\hat{d}_2$. Con base en estas observaciones, la ecuación 3.1 puede ser simplificada como:

$$TPC_p = p(P_{Rx}(\hat{d}_1) \geq P_{min}, P_{Rx}(\hat{d}_2) < P_{min}). \tag{3.2}$$

Por otro lado, la potencia recibida a una distancia $\hat{d}_i$ (i. e. la distancia entre el transmisor y el receptor) cuando se consideran las condiciones del canal inalámbrico, puede ser modelada como:

$$P_{Rx}(\hat{d}_i)\ \mathrm{dBm} = P_{Tx} + K - 10\gamma log(\frac{\hat{d}_i}{d_0}) + X_\sigma. \tag{3.3}$$

En donde P_{Tx} denota a la potencia de transmisión, el valor de la constante K depende de las caracterísitcas de la antena y la atenuación promedio del canal, d_0 es un valor de referencia relacionado con el "campo cercano" de la antena. El valor de K se refiere a la pérdida de la trayectoria en el espacio libre a la distancia $d_i < d_0$, cuyo valor es igual a $20log_{10}\frac{\lambda}{4\pi d_0}$ (típicamente el valor del parámetro d_0, en escenarios exteriores, se establece dentro del rango entre $1\,\mathrm{m}$ y $10\,\mathrm{m}$).γ es el exponente de pérdida por la trayectoria, y finalmente, X_σ es una variable aleatoria gaussiana con media igual a cero y desviación estándar igual a σ, esta variable representa el factor de desvanecimiento (*shadow-fading*). Cabe resaltar que el valor de la potencia de transmisión pueda ser afectada por factores tales como pérdidas en los conectores, la ganancia de la antenas de transmisión (para lo cual debe considerarse el PIRE) o variaciones propias de los componentes electrónicos. Sin embargo, en esta tesis dichos valores no serán considerados con el fin de simplificar la explicación de los modelos.

Tomando en cuenta la ecuación 3.3 y el modelo probabilístico 3.2, entonces TPC_p puede reescribirse en términos de la función de distribución de probabilidad conjunta como:

$$TPC_p = p(X \geq P_{min}, Y < P_{min}) = \int_{-\infty}^{P_{min}} \int_{P_{min}}^{\infty} \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{q(x,y)}{2}} dxdy, \tag{3.4}$$

en donde:

$$q(x,y) = \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]. \tag{3.5}$$

Donde X es una variable aleatoria gaussiana con media $\mu_x = P_{Tx} + K - 10\gamma log(\frac{\hat{d}_1}{d_0})$ y desviación estándar igual a σ_x. Asimismo, Y es una variable aleatoria gaussiana con media $\mu_y = P_{Tx} + K - 10\gamma log(\frac{\hat{d}_2}{d_0})$ y desviación estándar igual a σ_y. El parámetro ρ corresponde al coeficiente de correlación entre X e Y, estas variables representan la potencia recibida por los AP ubicados a las distancias $\hat{d}_1$ y $\hat{d}_2$, respectivamente.

En este trabajo, definimos a la potencia de transmisión óptima (P_{Tx}^{optima}) como el valor P_{Tx} que maximiza la probabilidad TPC_p, lo cual garantiza que solo el AP más cercano pueda decodificar la señal del nodo móvil. La figura 3.1 muestra varias curvas de TPC_p con respecto a P_{Tx} para diferentes valores de ρ. Para esta figura se consideraron los valores de $\hat{d}_1$ igual a $15\,\mathrm{m}$, $\hat{d}_2$ igual a $20\,\mathrm{m}$, $\sigma_x = \sigma_y = 6\,\mathrm{dB}$, $\lambda = 0.125\,\mathrm{m}$ y γ igual a 3.71. Se puede observar en esta figura que independientemente del valor de ρ, el valor de P_{Tx}^{optima} es igual a $-4\,\mathrm{dBm}$, en todos los casos. También, se puede observar

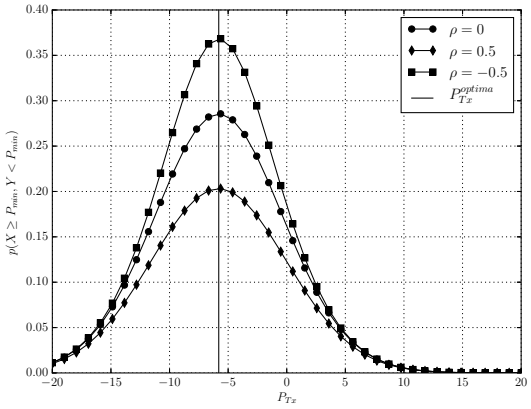


Figura 3.1: Potencia de transmisión óptima.

en la misma figura que el valor de P_{Tx}^{optima} es independiente del valor de ρ. Por consiguiente, podemos asumir el valor de $\rho = 0$ para obtener una expresión analítica de P_{Tx}^{optima}, mediante la expresión siguiente:

$$TPC_p = p(X \geq P_{min}, Y < P_{min}) = p(Y < P_{min})p(X \geq P_{min}). \tag{3.6}$$

Tomando en cuenta que X e Y son variables aleatorias gaussianas, el modelo anterior puede reescribirse de la siguiente manera [39]:

$$\begin{aligned} TPC_p &= (1 - Q(\frac{P_{min} - \mu_y}{\sigma_y}))Q(\frac{P_{min} - \mu_x}{\sigma_x}) \\ &= Q(-\frac{P_{min} - \mu_y}{\sigma_y})Q(\frac{P_{min} - \mu_x}{\sigma_x}). \end{aligned} \tag{3.7}$$

Para obtener el máximo de la ecuación 3.7 una alternativa es derivar la expresión TPC_p con respecto a P_{Tx} e igualarla a cero. Desafortunadamente, la derivada de la función Q no tiene una solución cerrada. Sin embargo, nos dimos cuenta que si se grafican los términos $Q(\frac{P_{min}-\mu_x}{\sigma_x})$ y $Q(-\frac{P_{min}-\mu_y}{\sigma_y})$ de la ecuación 3.7, el valor máximo de la función TPC_p corresponde con la intersección de ambas funciones de probabilidad, como se puede ver en la figura 3.2. En consecuencia, para obtener una expresión analítica que determine el valor de potencia de transmisión óptima, igualaremos ambas funciones Q como se muestra:

$$Q(-\frac{P_{min} - \mu_y}{\sigma_y}) = Q(\frac{P_{min} - \mu_x}{\sigma_x}) \tag{3.8}$$

Y por consiguiente:

$$-\frac{P_{min} - \mu_y}{\sigma_y} = \frac{P_{min} - \mu_x}{\sigma_x} \tag{3.9}$$

De esta manera, al sustituir las expresiones para μ_x y μ_y en la ecuación 3.9, y resolviendo para P_{Tx} obtenemos la siguiente expresión:

$$P_{Tx}^{optima} = P_{min} - K + 5\gamma log_{10}(\hat{d}_1 \cdot \hat{d}_2). \tag{3.10}$$

En la figura 3.2, se grafica TPC_p cuando el valor de $\hat{d}_1$ es igual a $30\,\mathrm{m}$, $\hat{d}_2$ es igual a $80\,\mathrm{m}$, $\sigma_x = \sigma_y = 6\,\mathrm{dB}$, $\lambda = 0.125\,\mathrm{m}$, γ es igual a 3.71 y $\rho = 0$. Se puede observar en esta figura que el valor de la potencia óptima es igual a $12.22\,\mathrm{dBm}$. En otras palabras, la probabilidad de que la transmisión del nodo móvil alcance solo al AP más cercano (i.e. el AP ubicado en la distancia $\hat{d}_1$) es igual $82\,\%$; sin embargo, hay un $18\,\%$ de probabilidad de que la señal del nodo móvil también alcance al AP ubicado a la distancia $\hat{d}_2$.

Por otro lado, aun cuando los nodos móviles pueden calcular la potencia de transmisión óptima mediante la ecuación 3.10, la implementación de dicha potencia en radios Wi-Fi puede resultar imposible. Esto se debe principalmente al hecho de que la mayoría de los radios que se encuentran en el mercado solo disponen de un intervalo limitado dentro del cual pueden ajustar su potencia de transmisión, típicamente entre 10 y $20\,\mathrm{dBm}$ [23]. Esto se agrava todavía más cuando consideramos que la potencia de transmisión óptima está en función de la densidad de los atacantes (esto afecta principalmente a las distancias $\hat{d}_1$ y $\hat{d}_2$), así como también al exponente de pérdida en la trayectoria γ (ver ecuación 3.10). En consecuencia, la combinación de estos dos factores puede, en algunos casos, generar valores de potencia de transmisión óptima muy por debajo de los niveles mínimos que un radio Wi-Fi puede alcanzar. Por ejemplo, en la referencia [23] los autores hacen un análisis de radios Wi-Fi con chips Atheros, ellos mostraron que la potencia de transmisión de dichos dispositivos puede ajustarse de manera estable entre 10 y $15\,\mathrm{dBm}$, y que dichos radios se vuelven inestables cuando la potencia de transmisión se encuentra por debajo de $5\,\mathrm{dBm}$.

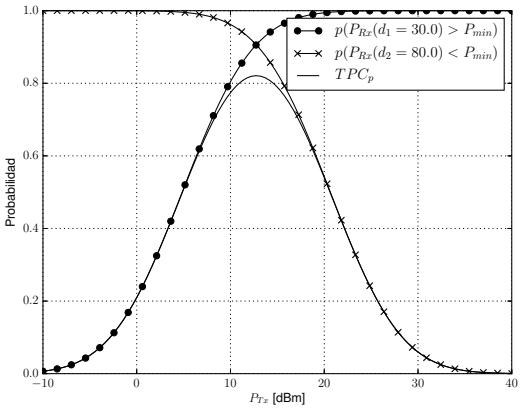


Figura 3.2: TPC_p para $\rho = 0$.

Sin embargo, aun cuando el valor estimado de potencia de transmisión se encuentra fuera del intervalo soportado por un radio Wi-Fi, en algunos casos, el usuario móvil puede utilizar un valor diferente de potencia de transmisión y aún conseguir un valor de TPC_p similar. Por ejemplo, en la figura 3.3 se muestran curvas de TPC_p para $\rho = 0$ y diferentes valores de σ , estas gráficas utilizan los mismos parámetros que la figura 3.2. Como se puede observar en esta figura, con un valor $P_{Tx}^{optima} = 0.11\,\mathrm{dBm}$ y un valor de $\sigma = 6\,\mathrm{dB}$, el valor de TPC_p corresponde a 0.96. Sin embargo, si ahora consideráramos una potencia de transmisión mayor, por ejemplo, $P_{Tx} = 7\,\mathrm{dBm}$, el valor de TPC_p disminuye a 0.84. En contraste, si se considera un valor de $\sigma = 2\,\mathrm{dB}$ y la misma $P_{Tx} = 7\,\mathrm{dBm}$, entonces el valor de $TPC_p = 0.99$. En otras palabras, la figura 3.3 muestra que en general, para valores de σ bajos (i.e. cercanos a dos), los nodos

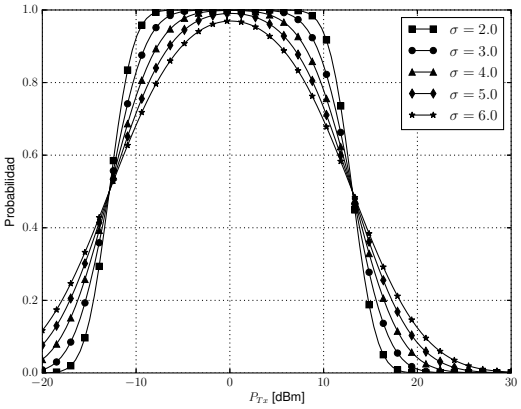
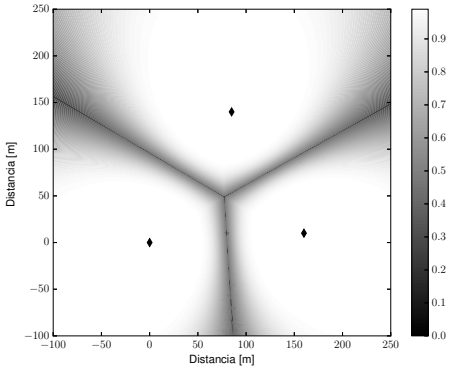


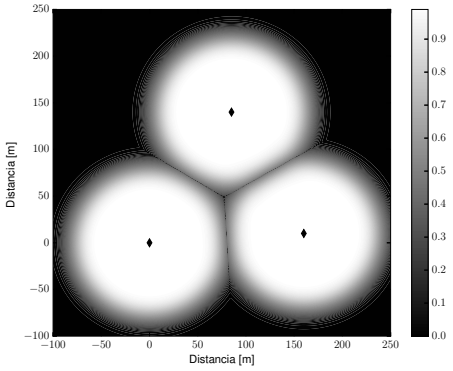
Figura 3.3: TPC_p para $P_{Tx} = 7\ dBm$ y diferentes valores de σ.

móviles tienen más posibilidades de seleccionar un valor de P_{Tx} dentro de su rango de operación y mantener un valor de TPC_p alto. Mientras que, para valores de σ cercanos a seis, seleccionar una potencia de transmisión diferente del valor óptimo siempre resulta en una disminución del valor de TPC_p.

Las figuras 3.4(a) y 3.5(a) muestran un mapa de valores TPC_p cuando un nodo móvil implementa el valor de potencia de transmisión óptima. Para estas figuras se utilizaron los parámetros descritos anteriormente en la figura 3.2. Los diamantes en estas figuras representan la posición de los AP, mientras que las áreas oscuras representan las posiciones donde el nodo móvil tiene una alta probabilidad de ser escuchado por más de un AP. Por el contrario, las áreas claras representan las posiciones en la cuales el nodo móvil tiene una alta probabilidad de ser escuchado y decodificado correctamente solo por el AP más cercano. Sin embargo, como mencionamos previamente, dado que los radios Wi-Fi solo pueden ajustar su potencia de transmisión en un intervalo limitado, las figuras 3.4(b) y 3.5(b) muestran el mapa de valores de TPC_p, en el caso cuando la potencia de transmisión fue limitada al intervalo de 10 a $20\,\mathrm{dBm}$. Es decir, si el valor de potencia de transmisión P_{Tx}^{optima} se encuentra por debajo de $10\,\mathrm{dBm}$, entonces se utilizó el valor de $10\,\mathrm{dBm}$. Por otro lado, si el valor de P_{Tx}^{optima} está por encima de $20\,\mathrm{dBm}$, entonces se utilizó el valor de $20\,\mathrm{dBm}$. Como se puede notar en estas figuras, las áreas oscuras incrementan su tamaño en comparación con las figuras 3.4(a) y 3.5(a), respectivamente. Finalmente, es importante notar que en canales más ruidosos (i.e. $\sigma \approx 6\,\mathrm{dB}$) el tamaño de las áreas oscuras se incrementa (figura 3.5(b)). Esto implica que los nodos móviles que implementan TPC tendrán menos lugares para engañar de manera efectiva a los atacantes. Adicionalmente, puede notarse que las áreas oscuras y claras en los mapas de TPC_p forman diagramas de Voronoi [40].

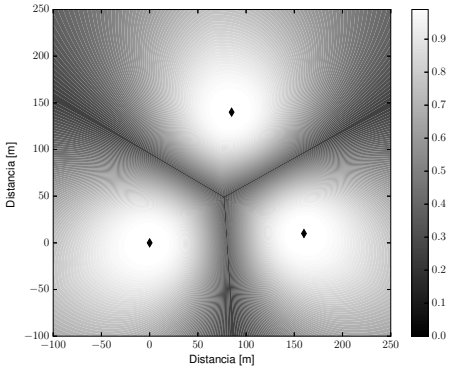


(a) TPC_p utilizando P_{Tx}^{Optima} y $\sigma = 2$.

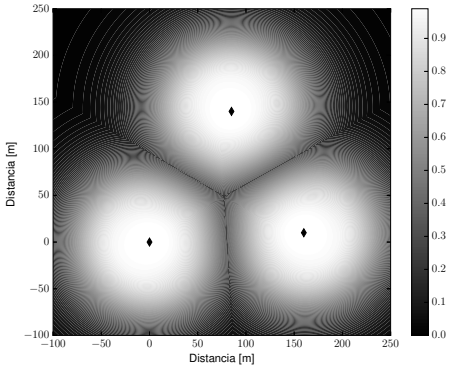


(b) TPC_p utilizando P_{Tx} en el intervalo entre 10 y $20\,\mathrm{dBm}$ y $\sigma = 2$.

Figura 3.4: Mapa de valores de TPC_p para $\sigma = 2\,\mathrm{dB}$ y $\rho = 0$.



(a) TPC_p utilizando P_{Tx}^{Optima} y $\sigma = 6$.



(b) TPC_p utilizando P_{Tx} en el intervalo entre 10 y $20\,\mathrm{dBm}$ y $\sigma = 6$.

Figura 3.5: Mapa de valores de TPC_p para $\sigma = 6\,\mathrm{dB}$ y $\rho = 0$.

El análisis previo con respecto a los mapas de TPC_p puede tener más aplicaciones. Por ejemplo, puede ayudar a un nodo móvil para buscar las rutas dentro del área de cobertura que tengan la mayor privacidad geográfica (p. ej. trazando rutas que crucen las aristas de Voronoi de forma perpendicular). Inclusive puede utilizarse para elegir una estrategia de ofuscación más acorde a la topología. Por el momento, se deja el desarrollo de un algoritmo para atravesar una área de cobertura con la mayor privacidad geográfica posible en el apartado de perspectivas de investigación.

3.3. Efectos de TPC en los algoritmos de localización

En esta sección se presenta la metodología utilizada para evaluar la eficacia de TPC como contramedida a los algoritmos de localización implementados por los atacantes en escenarios LE. Para ello, seleccionamos los algoritmos de localización más

representativos basados en estrategias *range-based*, así como en *range-free*. Posteriormente, se presenta el desarrollo de expresiones analíticas para calcular el error de estimación de los algoritmos de localización (i.e. la distancia entre la posición real de nodo móvil y la posición estimada). Finalmente, se presenta una breve discusión acerca de la relación entre el error de estimación y la cantidad de AP en un área determinada.

En general, podemos clasificar a los algoritmos de localización en dos grandes grupos, denominados: *range-based* y *range-free*. Las técnicas *range-based* suelen usar tres o más nodos fijos no colineales para estimar la posición de un nodo móvil mediante la solución de un problema geométrico. En contraste, los algoritmos de localización *range-free*, basan su operación en la conectividad de un nodo móvil, y estiman su ubicación mediante la solución de problemas heurísticos o de optimización. En la referencia [41], los autores comparan el funcionamiento de varios algoritmos de localización tanto del grupo *range-based*, como del grupo *range-free*. Este análisis concluye que, en general, los algoritmos *range-based* logran una mayor exactitud en la posición estimada en comparación con los algoritmos *range-free*. Sin embargo, esta exactitud viene con el costo de un mayor número de operaciones matemáticas. De acuerdo con los resultados obtenidos en [41], el algoritmo *range-based* más exacto es *circular positioning system* (CPS), mientras que el algoritmo *range-free* que logra la mayor exactitud es *weighted centroid localization* (WCL). Con base en estos resultados, se eligieron a estos dos algoritmos para evaluar la eficacia de TPC.

3.3.1. Algoritmo de localización CPS

La exactitud de los algoritmos de localización *range-based* está fuertemente ligada al número de AP que escuchan la señal del nodo móvil. La posición estimada en el algoritmo CPS, se obtiene a través de la intersección de las circunferencias cuyo centro es la posición de los AP, y cuyo radio es igual a la distancia estimada entre los AP y la posición actual del nodo móvil. En general, se puede expresar este problema como un sistema de ecuaciones:

$$(x - x_i)^2 + (y - y_i)^2 - d_i^2 = 0, \ \forall \ i = 1, 2, 3, ... \tag{3.11}$$

En el cual, cada ecuación describe una circunferencia con centro en el i-ésimo AP, cuyas coordenadas son (x_i, y_i). El radio de dicha circunferencia corresponde a d_i, y los parámetros (x, y), representan la posición estimada del nodo móvil. Para resolver este sistema de ecuaciones se deben considerar tres posibles casos:

- En el caso cuando el nodo móvil esté al alcance de tres o más nodos fijos, el sistema de ecuaciones tendrá una única solución. Esta solución representa la posición estimada del nodo móvil. A este caso lo denominamos escenario 3-AP.

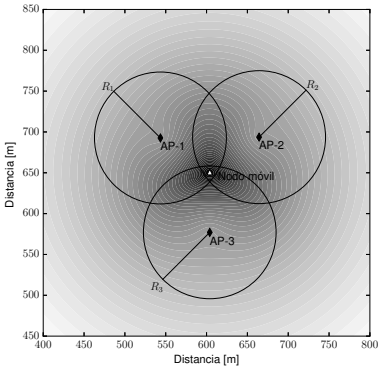
- En el caso cuando el nodo móvil esté al alcance de dos nodos fijos, existen dos puntos que dan solución al sistema de ecuaciones. Un punto corresponde a la posición estimada del nodo móvil, mientras que el otro punto corresponde a una solución matemática. Sin embargo, sin ninguna otra fuente de información, el algoritmo de localización no puede identificar cuál de los dos puntos corresponde a la posición actual del nodo móvil. A este caso lo denominamos escenario 2-AP.

- Por último, en el caso cuando el nodo móvil se encuentra solamente al alcance de un nodo fijo, existe una infinidad de puntos que pueden satisfacer el sistema de ecuaciones. En específico, estos puntos pertenecen a una circunferencia cuyo centro es la posición del AP, y el radio es la distancia estimada entre el AP y el nodo móvil. Este caso corresponde con el escenario 1-AP mencionado en la sección 3.1.

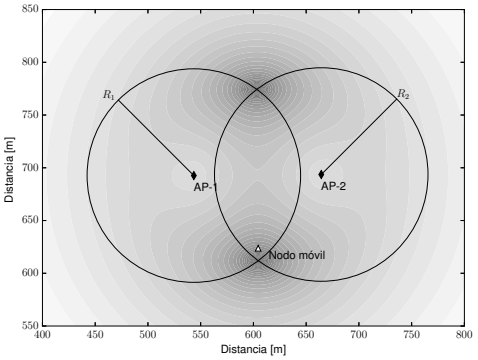
Por otro lado, debido a las condiciones cambiantes del medio inalámbrico, la distancia d_i estimada entre la posición del nodo móvil y algún AP puede presentar variaciones con respecto al valor real (ver ecuación 3.3). Esta situación provoca que en algunos casos el sistema de ecuaciones no tenga solución, ya que ninguna solución satisface las ecuaciones simultáneamente. Con el fin de resolver este problema, algunos trabajos han propuesto metodologías para obtener una solución aproximada. Por ejemplo, en la referencia [42], los autores reemplazan la intersección de circunferencias por la intersección de lineas rectas. En la referencia [41], se presenta un enfoque en el cual los autores formulan el sistema de ecuaciones como un problema de mínimos cuadrados, como se muestra a continuación:

$$f(x,y) = \sum_{i=0}^{m} (\sqrt{(x-x_i)^2 + (y-y_i)^2} - d_i)^2. \tag{3.12}$$

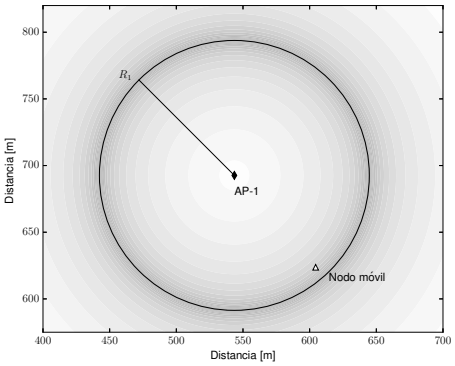
Donde el mínimo global de la función $f(x,y)$ representa la posición estimada del nodo móvil. Las figuras 3.6(a), 3.6(b) y 3.6(c), muestran los tres escenarios de localización mencionados previamente: 3-AP, 2-AP y 1-AP, respectivamente. En estas figuras, los diamantes negros representan las posiciones de los AP, las áreas oscuras ilustran las regiones de mayor probabilidad donde se encuentra el nodo móvil, mientras que la posición actual del nodo móvil se muestra con un triángulo blanco. Como se mencionó anteriormente, sin otra fuente de información, CPS es incapaz de distinguir entre la solución matemática y la posición actual del nodo móvil ver figura 3.6(b). Más aún, en el escenario 1-AP (figura 3.6(c)) CPS tampoco es capaz de distinguir la posición actual del nodo móvil en un infinito de posibilidades. Esta deficiencia de los algoritmos de localización es aprovechada por TPC, dado que en esta técnica el objetivo consiste en estimar la potencia de transmisión óptima que le garantice a un nodo móvil que solo un AP pueda escuchar su señal inalámbrica, es decir estar en el escenario 1-AP.



(a) Escenario 3-AP.



(b) Escenario 2-AP.



(c) Escenario 1-AP.

Figura 3.6: Funcionamiento del algoritmo CPS.

3.3.2. Error de estimación del algoritmo CPS

Como se mencionó previamente, el objetivo de TPC es lograr que los algoritmos de localización se encuentren en el escenario 1-AP el mayor tiempo posible. De esta manera, para evaluar la efectividad de TPC como técnica de privacidad geográfica, es necesario desarrollar expresiones analíticas que permitan calcular el error generado por CPS para los escenarios 1-AP, 2-AP y 3-AP.

En cuanto al escenario 1-AP, el algoritmo de localización CPS no puede distinguir cuál de todas las posibles soluciones corresponde con la posición actual del nodo móvil. Por consiguiente, para considerar esta situación específica, definimos el error de estimación como el promedio de las distancias entre la posición actual del nodo móvil y cada una de las posibles soluciones. La figura 3.7 muestra un caso particular, en el cual, la posición actual del nodo móvil está representada por un círculo negro, mientras la posición estimada está dibujada con un triángulo negro. En esta figura, el error de estimación (ver línea discontinua) es una función del ángulo θ y su valor puede calcularse usando la siguiente expresión:

$$\text{1-AP}_{error} = \frac{1}{2\pi}\int_0^{2\pi} \sqrt{d_{real}^2 + d_{estimada}^2 - 2d_{real}d_{estimada}cos(\theta)}d\theta \tag{3.13}$$

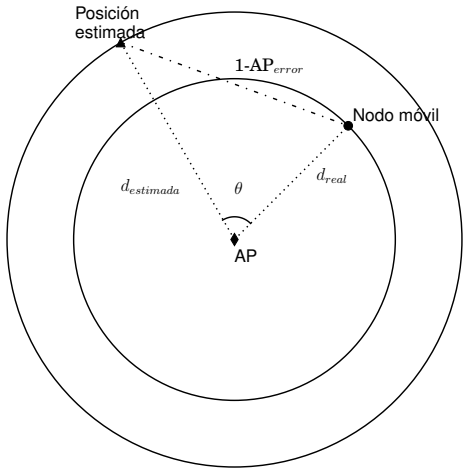


Figura 3.7: Error de estimación del algoritmo CPS en escenario 1-AP.

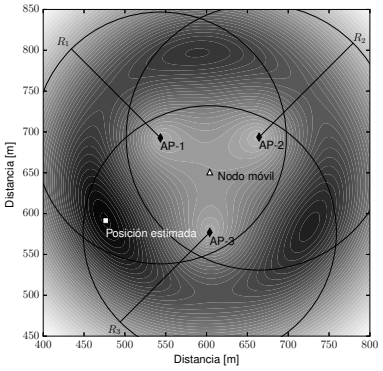
En esta expresión, el término d_{real} representa la distancia real entre el nodo móvil y el AP. El término $d_{estimada}$ denota la distancia estimada del nodo móvil, según el AP. θ es el ángulo formado entre la posición actual y la estimada del nodo móvil. De esta manera, para obtener el promedio de todas las distancias se calcula la integral de 0 a 2π con respecto θ dividida por 2π.

Por otro lado, para el escenario 2-AP existen dos soluciones, por lo que el error de estimación puede obtenerse mediante un promedio de las distancias euclidianas entre la posición actual del nodo móvil y las dos posiciones estimadas. Esta expresión se muestra a continuación:

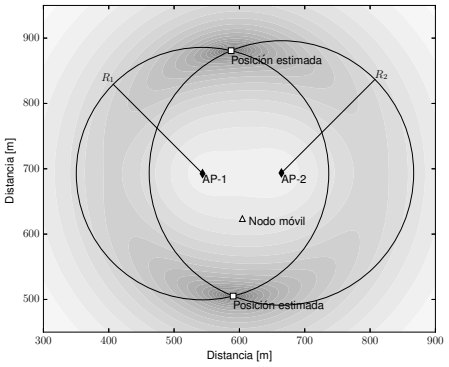
$$\text{2-AP}_{error} = \frac{1}{2}\sum_{i=1}^{2} d(P_{mn}, P^{i}_{estimated}). \tag{3.14}$$

En esta expresión, el término P_{mn} representa la posición actual del nodo móvil, mientras que $P^{i}_{estimated}$ representa la i-ésima posición estimada obtenida del sistema de ecuaciones. $d(P_{mn}, P^{i}_{estimated})$ es la distancia euclidiana entre la posición real del nodo móvil y la i-ésima posición estimada. Finalmente, el error de estimación para el escenario 3-AP (3-AP$_{error}$) se obtiene mediante la distancia euclidiana entre la posición real del nodo móvil y la única solución del sistema de ecuaciones.

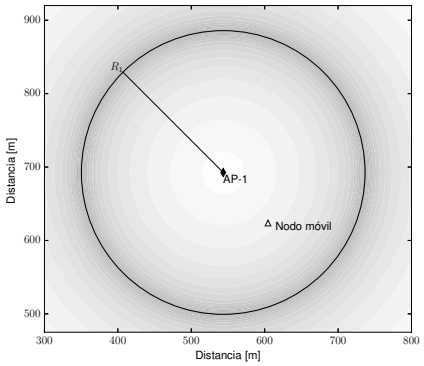
Aun cuando CPS es uno de los algoritmos de localización más exactos, su error de estimación depende en gran medida del número de AP considerados en la estimación, así como la distancia d_i reportada por cada AP. No obstante, aun cuando el número de AP sea igual o mayor a tres, si el conjunto de distancias estimadas por los AP difieren considerablemente de las distancias reales, la exactitud del algoritmo se reduce. Por ejemplo, en la figura 3.8(a) se ilustra un escenario 3-AP, en el cual, debido al error en las distancias estimadas por los AP, no existe un punto de intersección que satisfaga las tres circunferencias simultáneamente. En este caso, el mínimo global encontrado por el algoritmo de mínimos cuadrados está en una posición lejana en comparación con la posición actual del nodo móvil (representada por un cuadrado blanco). Las figuras 3.8(b) y 3.8(c) ilustran la misma situación para los escenarios 2-AP y 1-AP, respectivamente.



(a) Escenario 3-AP.



(b) Escenario 2-AP.



(c) Escenario 1-AP.

Figura 3.8: Funcionamiento del algoritmo CPS cuando nodos móviles implementan TPC.

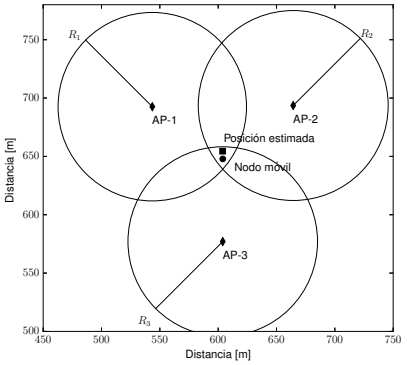
3.3.3. Algoritmo de localización WCL

El algoritmo WCL estima la posición de un nodo móvil mediante un promedio ponderado de todas las posiciones de los AP que escuchan al nodo móvil. Para ponderar cada coordenada se utiliza un peso cuyo valor se obtiene de medir un parámetro físico como: RSSI, ToA, LQI o TDoA. La ecuación 3.15 muestra las expresiones para obtener la posición estimada $(\hat{x}, \hat{y})$ de un nodo móvil.

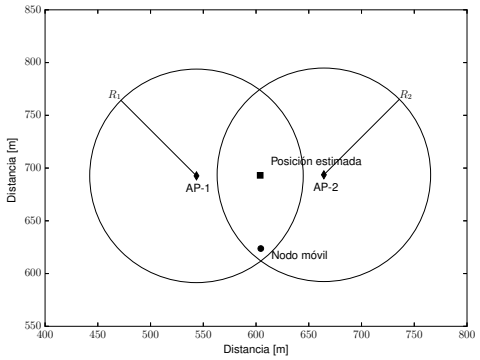
$$\begin{aligned} \hat{x} &= \frac{\sum_{i=1}^{m} w_i x_i}{\sum_{i=1}^{m} w_i}, \\ \hat{y} &= \frac{\sum_{i=1}^{m} w_i y_i}{\sum_{i=1}^{m} w_i}. \end{aligned} \tag{3.15}$$

En donde las coordenadas (x_i, y_i) corresponden a las coordenadas del i-ésimo AP que escucha las transmisiones del nodo móvil. m es el número total de AP que escuchan al nodo móvil. Cabe resaltar que, en el caso cuando solo hay un AP (i.e. el escenario 1-AP), el valor del peso w_1 se cancela. Por consiguiente, la posición estimada del nodo móvil corresponde a la posición del único AP. La figura 3.9(c) ilustra este caso, donde la posición estimada del nodo móvil está representada por un cuadrado negro,

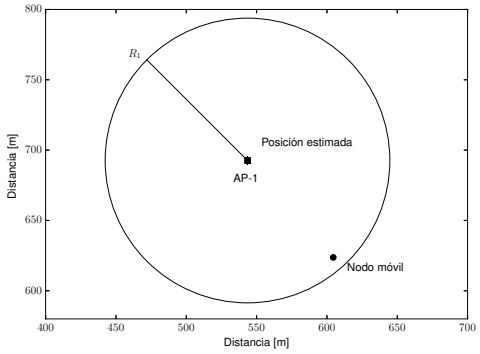
mientras que la posición actual del nodo móvil está representada como un círculo negro. Las figuras 3.9(a) y 3.9(b) muestran la misma situación para los escenarios 2-AP y 3-AP, respectivamente.



(a) Escenario 3-AP.



(b) Escenario 2-AP.



(c) Escenario 1-AP.

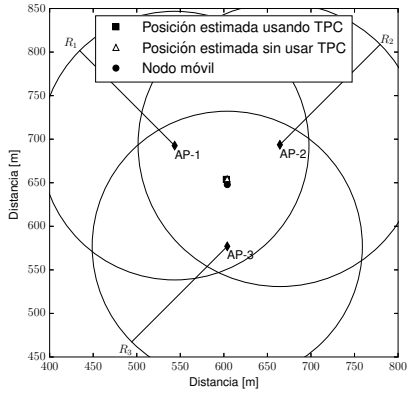
Figura 3.9: Funcionamiento del algoritmo WCL.

3.3.4. Error de estimación del algoritmo WCL

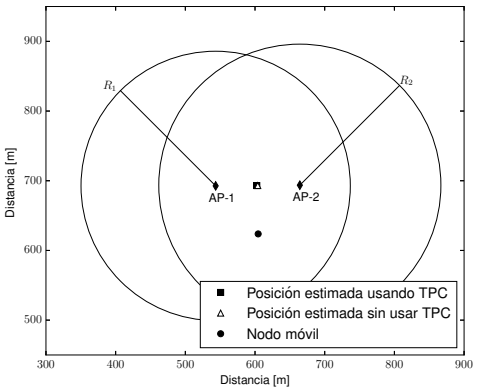
Contrario al funcionamiento de CPS donde se puede tener una, dos o infinitas soluciones, WCL siempre obtiene una posición estimada, independientemente del número de AP que escuchan la señal del nodo móvil. Por consiguiente, el error de estimación de WCL se calcula como la distancia euclidiana entre la posición actual del nodo móvil y la posición estimada.

Cabe resaltar que la posición estimada en WCL depende exclusivamente del cociente entre la suma ponderada de las coordenadas de los AP y la suma de los pesos w_i (ver ecuación 3.15). Esta característica provoca que la posición estimada sea menos susceptible a ser afectada por TPC. Esto se debe principalmente al hecho de que las variaciones en la potencia de transmisión afectan de igual manera a todos los AP que escuchan al nodo móvil, y por consiguiente el cociente reduce este efecto. Este comportamiento es ilustrado en la figura 3.10 para los escenarios 3-AP, 2-AP y 1-AP; se puede observar en esta figura que la posición estimada del nodo móvil cuando este no implementa TPC (cuadrado negro) comparada con la posición estimada cuando

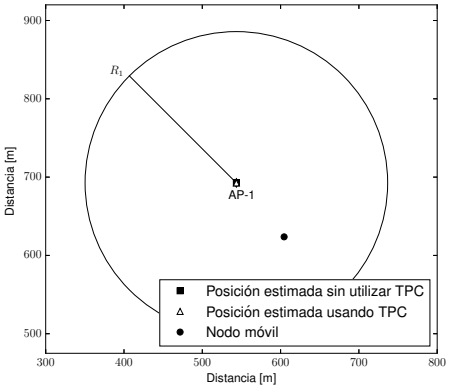
este implementa TPC (triángulo blanco), son muy similares.



(a) Escenario 3-AP.



(b) Escenario 2-AP.



(c) Escenario 1-AP.

Figura 3.10: Funcionamiento del algoritmo WCL cuando nodos móviles implementan TPC.

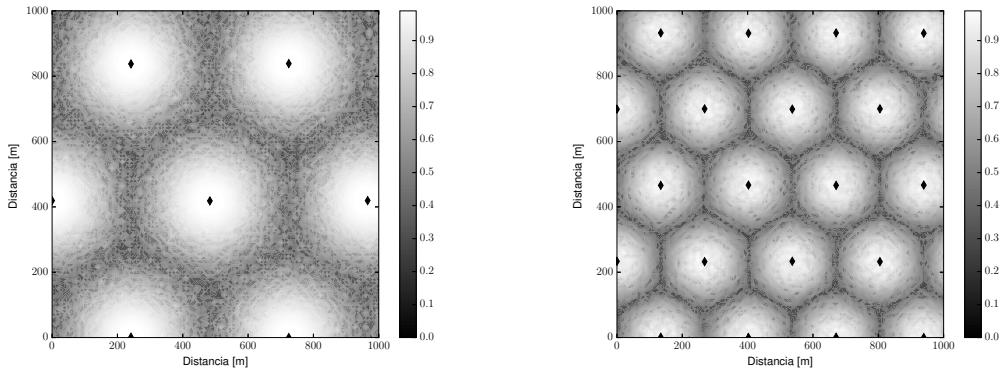
3.4. Simulaciones y pruebas en exteriores

En esta sección se presenta la evaluación de la eficacia de TPC como técnica de privacidad geográfica. En primer lugar se presenta un conjunto de simulaciones con el fin de evaluar el error de estimación de CPS y WCL bajo diversas condiciones del medio inalámbrico, así como diferentes densidades de atacantes. En segundo lugar se presentan las pruebas realizadas en un entorno exterior para corroborar los resultados de las simulaciones.

3.4.1. Simulaciones

Con el objetivo de evaluar la eficacia de TPC como contramedida a los algoritmos de localización CPS y WCL, en esta tesis se diseñó un conjunto de simulaciones que consideran diversas condiciones del medio inalámbrico así como diferentes densidades de atacantes. Estas simulaciones se realizaron en un simulador implementado en el

lenguaje de programación Python. La región considerada corresponde a un cuadrado de $1000\,\mathrm{m} \times 1000\,\mathrm{m}$. En esta región se ubicaron los AP siguiendo un patrón triangular propuesto en la referencia [43]. De esta manera, la posición de cada AP corresponde con el vértice de un triángulo equilátero cuyo lado tiene una longitud igual a L. En consecuencia, al variar la longitud L, el número de AP dentro de la región cambia gradualmente. En específico, en esta tesis se define a la densidad de los AP como el cociente L/R, en donde R corresponde al máximo alcance de comunicación entre un transmisor y un receptor, y L es la distancia euclidiana entre dos AP consecutivos. El valor de R en estas simulaciones es expresado en función de los siguientes parámetros: pérdida por la trayectoria γ, el factor de desvanecimiento σ, la potencia máxima de transmisión $P_{Tx} = 20\,\mathrm{dBm}$ y la probabilidad de decodificar correctamente un paquete de datos. En particular, el valor seleccionado de esta probabilidad para las simulaciones fue igual o mayor que $50\,\%$. Las figuras 3.11(a) y 3.11(b) muestran valores de TPC_p en un escenario de baja densidad (i.e. $L/R = 2.25$) y un escenario de alta densidad (i.e. $L/R = 1.25$), respectivamente. En estas figuras, las regiones oscuras representan las zonas en las cuales el nodo móvil puede ser escuchado por más de un AP. Asimismo, las regiones claras representan las zonas en las cuales el nodo móvil tiene una alta probabilidad de alcanzar solamente al AP más cercano.



(a) TPC_p para un escenario con baja densidad de AP ($L/R = 2.25$).

(b) TPC_p para un escenario con alta densidad de AP ($L/R = 1.25$).

Figura 3.11: Mapa de valores de TPC_p para $\gamma = 3$ y $\sigma = 6\,\mathrm{dB}$.

Con el fin de calcular los valores TPC_p y P_{Tx}^{optima}, el modelo probabilístico presentado en la sección 3.2 requiere como variable de entrada el valor de la distancia estimada entre un transmisor y un receptor. En la literatura se han propuesto diversas técnicas para calcular dicha distancia. Por ejemplo, en las referencias [44, 45] los autores convierten el valor promedio de mediciones de RSSI a distancia utilizando el modelo de propagación *path-loss*. Por otro lado, en las referencias [46, 47], los autores proponen una técnica más exacta comparada con [44, 45], la cual utiliza estimadores probabilísticos. Sin embargo, estos estimadores no consideran la variación de la potencia de transmisión. En la referencia [48], los autores proponen un estimador de distancia, el cual considera las condiciones del medio inalámbrico y requiere solamente una muestra de mediciones RSSI. Por esta razón, en esta sección utilizaremos este estimador de distancia, sin embargo, cualquier metodología que permita relacionar la potencia de transmisión con la distancia entre un receptor y un transmisor puede ser utilizada.

Como se mencionó en la sección 3.2, el primer paso de la técnica TPC consiste en que el nodo móvil estime la distancia entre su posición y todos los AP cercanos. Para conseguir este objetivo, el nodo móvil obtiene de cada AP una muestra de mediciones de potencia recibida P_{Rx}. Estas muestras se generan a través de la ecuación 3.3, donde $\hat{d}_i$ es la distancia euclidiana entre el nodo móvil y el i-ésimo AP, el parámetro P_{Tx} es igual a $20\,\mathrm{dBm}$, $K = -40\,\mathrm{dB}$, $\gamma = 3$, mientras que X_σ es una variable aleatoria con media cero y desviación estándar $\sigma = 6\,\mathrm{dB}$. De esta forma, el nodo móvil puede convertir un conjunto de muestras de P_{Rx} a distancia usando la siguiente expresión:

$$d_{i,j} = 10^{\frac{P_{Tx}+K-P_{Rx_j}}{10\gamma}}, \tag{3.16}$$

en donde $d_{i,j}$ representa la j-ésima muestra del i-ésimo AP. Finalmente, la distancia estimada entre el nodo móvil y el i-ésimo AP puede calcularse mediante la ecuación 3.17 [48], en donde $\bar{d}_i$ es la media muestral y $\bar{s}_i$ es la desviación estándar.

$$\hat{d}_i = \sqrt{\frac{\bar{d}_i^4}{\bar{d}_i^2 + \bar{s}_i^2}}. \tag{3.17}$$

Una vez que el nodo móvil haya obtenido el conjunto de distancias estimadas entre su ubicación y la de cada AP vecino, dicho conjunto debe ordenarse de forma ascendente para obtener las distancias $\hat{d}_1$ y $\hat{d}_2$ (ver Sección 3.2). De esta manera, el nodo móvil puede calcular la potencia de transmisión óptima utilizando la ecuación 3.10, y TPC_p mediante la ecuación 3.4.

Con el fin de evaluar la eficacia de TPC, tres casos deben ser considerados:

- El nodo móvil transmite señales inalámbricas utilizando la potencia de transmisión máxima, $20\ dBm$. Haremos referencia a este caso como No-TPC.
- El nodo móvil transmite sus señales inalámbricas utilizando la potencia de transmisión óptima, calculada a partir de la ecuación 3.10. Haremos referencia a este caso como TPC óptimo (OTPC).
- El nodo móvil calcula el valor de su potencia de transmisión óptima, pero transmite utilizando la potencia de transmisión acotada dentro del intervalo de 10 a $20\,\mathrm{dBm}$. Haremos referencia a este caso como TPC limitado (LTPC).

Cada experimento fue ejecutado 50 veces con el fin de obtener valores promedio. Para estos experimentos, definimos arbitrariamente el valor de γ y σ igual a 3 y $6\,\mathrm{dB}$, respectivamente. El valor de R corresponde a $215\,\mathrm{m}$, dicho valor fue calculado utilizando la ecuación 3.18[39] y fijando una probabilidad para decodificar un paquete correctamente igual o mayor a $50\,\%$. La figura 3.12(a) muestra el error de estimación promedio, obtenido por CPS, en la región estudiada, y considerando diferentes condiciones de densidad de AP. Se puede ver en esta figura que el error de estimación promedio para el caso No-TPC disminuye mientras la densidad de AP aumenta. Por ejemplo, para el valor de densidad 0.5 se tiene un error de estimación promedio igual a $4\,\mathrm{m}$, mientras que para una densidad de 3.5 el error de estimación promedio corresponde a $280\,\mathrm{m}$. Por el contrario, en OTPC y LTPC el error de estimación promedio muestra una tendencia creciente cuando se incrementa la densidad de atacantes. En específico, este comportamiento es más notable para valores de densidad menores o iguales a uno. Este resultado comprueba el efecto de TPC sobre el algoritmo CPS descrito en el análisis teórico presentado en la sección 3.3.2. En esta figura también

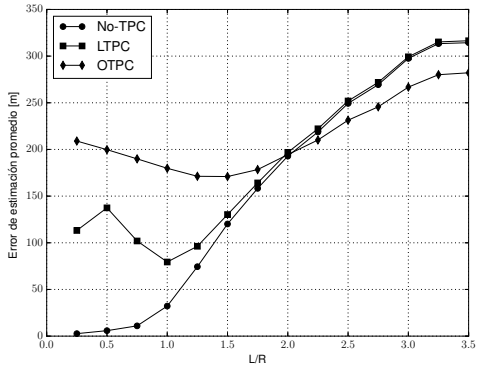
se puede ver que el error de estimación promedio para el caso LTPC es menor al error de estimación para el caso OTPC. Por ejemplo, cuando la densidad es igual a 0.25, el error de estimación promedio para LTPC es de alrededor de $100\,\mathrm{m}$ más pequeño que el valor para OTPC. Esto se debe principalmente al efecto de acotar la potencia de transmisión óptima en el intervalo de 10 a $20\,\mathrm{dBm}$. Este efecto es más notorio para los casos cuando L/R es igual a 0.5 y 0.25, en los cuales el error de estimación promedio disminuye a causa de la limitación en la potencia de transmisión (ver figura 3.12(c) en los mismos valores).

$$p(P_{Rx}(\hat{d}_i) \leq P_{min}) = 1 - Q(\frac{P_{min} - (P_{Tx} + K - 10\gamma log_{10}(\frac{\hat{d}_i}{d_0}))}{\sigma_{\psi_{dB}}}). \tag{3.18}$$

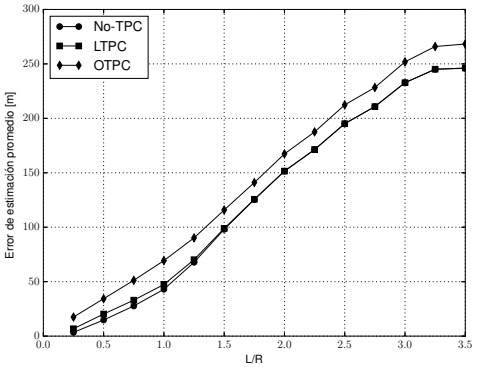
Por otro lado, en la figura 3.12(c) se muestra la potencia de transmisión promedio utilizada por el nodo móvil en los diferentes casos de densidad de AP. Como puede verse en esta figura, par el caso OTPC en escenarios con baja densidad de AP (i.e., valores de $L/R \geq 2.0$) la potencia de transmisión promedio fue mayor a $20\,\mathrm{dBm}$, la cual es comúnmente la potencia máxima de transmisión de radios Wi-Fi. En caso contrario, para el caso OTPC en escenarios con una alta densidad de AP (i.e. valores de $L/R \leq 1.0$) la potencia de transmisión promedio está por debajo de $10\,\mathrm{dBm}$. También puede observarse que la potencia de transmisión promedio para el caso LTPC está entre 10 y $20\,\mathrm{dBm}$.

Como se mencionó anteriormente, en CPS, el error de estimación está directamente relacionado con el número de AP que escuchan la señal del nodo móvil, y el error máximo corresponde al escenario 1-AP. Este comportamiento puede verse claramente en la figura 3.12(d), en la cual el número promedio de AP tiene un valor cercano a uno para el caso OTPC. Sin embargo, limitar la potencia de transmisión tiene un efecto negativo en el número promedio de AP que detectan el nodo móvil (ver línea con cuadrados). Este efecto es más notorio en escenarios de alta densidad (i.e. valores de $L/R \leq 1.0$), en los cuales el número de AP que escuchan al nodo móvil se incrementa. Por ejemplo, para la densidad igual a 0.25, el número de AP promedio para el caso de LTPC es igual a 10. Este resultado indica que utilizar radios Wi-Fi con potencia limitada disminuye la capacidad de TPC para reducir el número de AP. Cabe resaltar, que para este ejemplo, la potencia de transmisión óptima debía ser $-8\,\mathrm{dBm}$ (ver figura 3.12(c)), sin embargo, este valor de potencia de transmisión está muy por debajo del valor mínimo disponible de un radio Wi-Fi.

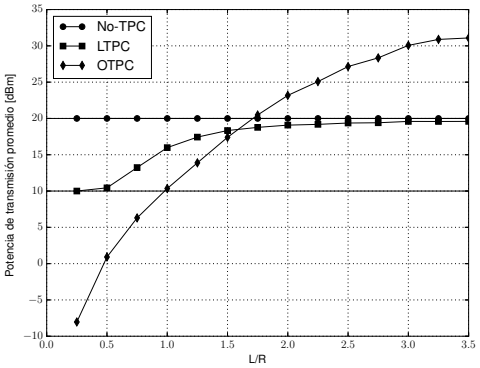
Los resultados de las simulaciones también corroboraron el hecho de que el algoritmo WCL es menos sensible a los efectos de TPC que el algoritmo CPS, como se explicó en la sección 3.3.4. Este comportamiento puede verse en la figura 3.12(b), en la cual el error de estimación, tanto para el caso OTPC como para LTPC, resultaron similares al error encontrado en el caso No-TPC. Adicionalmente, se calculó el error *root mean square* (RMS) entre los casos OTPC y No-TPC el cual resultó igual a $19.44\,\mathrm{m}$, mientras que el error RMS para el caso de LTPC comparado con No-TPC resultó de $2.60\,\mathrm{m}$. Estos resultados muestran que TPC no es capaz de garantizar privacidad geográfica cuando los atacantes implementan el algoritmo WCL.



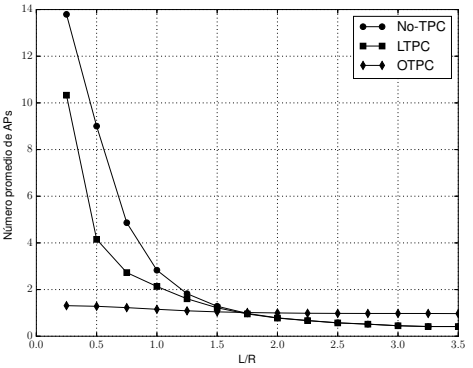
(a) Error de estimación de CPS.



(b) Error de estimación de WCL.



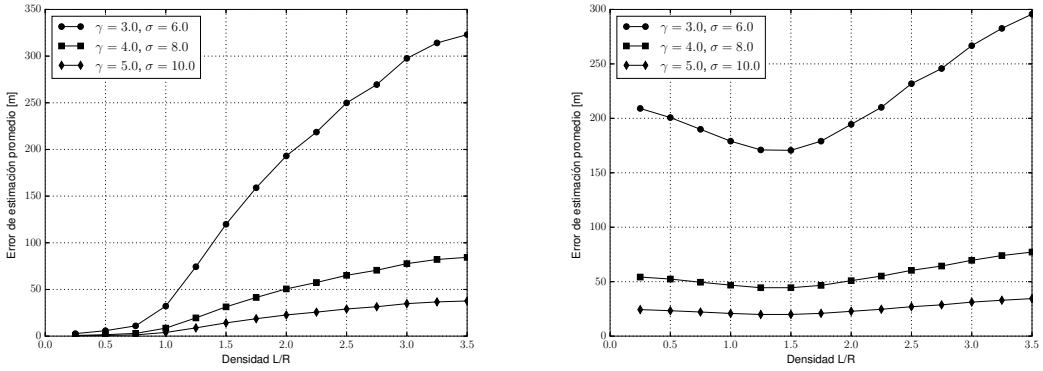
(c) Potencia de transmisión.



(d) Número promedio de AP.

Figura 3.12: Resultados de simulación para $\gamma = 3$ y $\sigma = 6$.

Finalmente, evaluamos la eficiencia de TPC como contramedia al algoritmo de localización CPS bajo diferentes condiciones del canal inalámbrico (i.e. diferentes combinaciones de γ y σ). Los resultados de la simulación muestran que aun cuando un nodo móvil no implemente TPC, el error de estimación inherente a CPS está fuertemente relacionado con las condiciones del canal inalámbrico. Este comportamiento se muestra en la figura 3.13(a), en la cual se puede notar como el error de estimación para el caso No-TPC decrece mientras los valores de γ, σ y densidad de AP se incrementan. También se puede notar en esta figura que incluso en el escenario 1-AP ($L/R = 3.5$), el error de estimación promedio cuando $\gamma = 5.0$ y $\sigma = 10.0\,\mathrm{dB}$ es aproximadamente $280\,\mathrm{m}$ menor que el error de estimación promedio para el caso cuando $\gamma = 3.0$ y $\sigma = 6.0\,\mathrm{dB}$. Más aún, cuando el nodo móvil implementa TPC, el error de estimación promedio está fuertemente relacionado con las condiciones del canal inalámbrico. Esto se puede ver ejemplificado en la figura 3.13(b), en la cual, el error de estimación cuando $\gamma = 5.0$ y $\sigma = 10.0\,\mathrm{dB}$ es mucho menor que el error de estimación cuando $\gamma = 3.0$ y $\sigma = 6.0\,\mathrm{dB}$.



(a) Caso No-TPC. (b) Caso OTPC.

Figura 3.13: Error de estimación promedio para diferentes valores de γ y σ.

3.4.2. Pruebas en entornos exteriores

Con el fin de evaluar la eficacia de TPC bajo condiciones realistas, realizamos una prueba en un entorno exterior utilizando radios Wi-Fi bajo el estándar IEEE 802.11n, el cual consta de árboles, calles, autos y edificios que provocan que la señal inalámbrica tenga interferencia y pérdida de energía. Este experimento consistió en desplazar un nodo móvil a través del campus de Ciudad Universitaria, mientras que un conjunto de AP recolectaban mediciones de RSSI del nodo móvil. La región considerada en este caso corresponde a un rectángulo de $400\,\mathrm{m} \times 200\,\mathrm{m}$, como se muestra en la figura 3.14. Adicionalmente, se consideró el origen del sistema de coordenadas en la esquina superior izquierda. En esta figura, los círculos negros representan los puntos por los cuales pasó el nodo móvil, mientras que los triángulos indican la ubicación de los AP durante la prueba. En esta figura también se puede ver que, a lo largo de la ruta, el nodo móvil encontró regiones de cobertura en donde uno, dos, tres o más AP escucharon sus transmisiones. Antes de llevar a cabo este experimento, se recolectaron un conjunto de mediciones RSSI provenientes del nodo móvil a diferentes distancias, con el fin de caracterizar el canal inalámbrico. Dichas mediciones se obtuvieron utilizando un radio Atheros a $2.4\,\mathrm{GHz}$ bajo el estándar IEEE 802.11n, de acuerdo al fabricante la sensibilidad mínima de nuestro dispositivo es igual a $-90\,\mathrm{dBm}$. De acuerdo con estas mediciones, el alcance máximo de comunicación fue $63\,\mathrm{m}$, $\gamma = 4.19$, $K = -40\,\mathrm{dB}$, $\lambda = 0.125\,\mathrm{m}$, y $\sigma = 5.98\,\mathrm{dB}$.

Para cada una de los posiciones en la ruta establecida (ver figura 3.14), el experimento constó de dos pasos. En el primer paso, el nodo móvil transmitió cien paquetes usando la potencia de transmisión máxima (en nuestro caso, $P_{Tx} = 20\,\mathrm{dBm}$). Simultáneamente, los AP que escucharon dicha transmisión recolectaron las mediciones de RSSI para llevar a cabo la localización del nodo móvil (i.e. el caso No-TPC). En el segundo paso, el nodo móvil transmitió cien paquetes utilizando la potencia de transmisión óptima, pero acotada al intervalo entre 10 y $20\,\mathrm{dBm}$. De igual manera, los AP que escucharon la transmisión recolectaron mediciones de RSSI para estimar la posición del nodo móvil (i.e. el caso LTPC). Este proceso se repitió 30 veces para obtener valores promedio. Los resultados obtenidos de esta prueba se muestran en la tabla 3.1. La primera columna en esta tabla indica la posición en la cual se llevaron a cabo las mediciones (ver figura 3.14). La segunda columna muestra el número promedio de

Tabla 3.1: Pruebas en escenarios exteriores

Pos.	**No-TPC**				**LTPC**			
	AP	P_{Tx} [dBm]	**Error de estimación**		**AP**	P_{Tx} [dBm]	**Error de estimación**	
			CPS [m]	**WCL [m]**			**CPS [m]**	**WCL [m]**
l_1	3.0	20.0	2.4	5.7	2.4	10.0	31.5	8.4
l_2	2.9	20.0	7.5	8.9	2.2	10.0	34.3	12.1
l_3	2.9	20.0	1.2	18.2	2.0	10.0	42.1	18.5
l_4	2.7	20.0	10.8	29.9	1.4	10.0	62.1	31.3
l_5	2.5	20.0	21.6	42.9	0.4	10.0	113.4	55.7
l_6	2.3	20.0	32.7	41.1	0.4	11.0	101.2	49.3
l_7	1.9	20.0	11.6	23.8	0.5	12.0	93.7	51.3
l_8	1.4	20.0	37.7	31.5	0.7	12.0	76.9	42.1
l_9	1.2	20.0	26.3	21.2	1.0	10.0	48.2	23.7
l_{10}	1.2	20.0	16.9	13.6	1.0	10.0	30.8	15.1
l_{11}	1.1	20.0	24.4	19.6	1.0	10.0	39.8	19.6
l_{12}	1.1	20.0	37.5	29.9	0.8	10.0	66.0	32.4
l_{13}	1.8	20.0	20.3	18.3	0.6	10.0	74.9	36.8
l_{14}	2.1	20.0	11.2	13.7	1.1	10.0	40.7	16.2
l_{15}	2.1	20.0	11.1	14.4	1.0	10.0	24.2	11.7
l_{16}	2.6	20.0	20.7	18.8	1.1	10.0	27.4	13.6
l_{17}	2.9	20.0	32.0	23.1	1.1	10.0	45.2	22.1
l_{18}	3.2	20.0	19.9	14.2	2.1	10.0	32.1	14.1
l_{19}	2.7	20.0	10.1	10.5	2.0	10.0	25.3	10.5
l_{20}	2.8	20.0	7.9	13.4	1.3	10.0	36.2	17.1
l_{21}	3.8	20.0	2.6	19.7	1.4	10.0	54.7	24.3
l_{22}	3.8	20.0	5.4	21.9	1.2	10.0	67.8	31.0
l_{23}	3.7	20.0	1.1	23.9	1.1	10.0	81.1	39.0
	2.42*	20*	16.21*	20.79*	1.20*	10.21*	54.33*	25.90*

*Valores promedio de toda la ruta

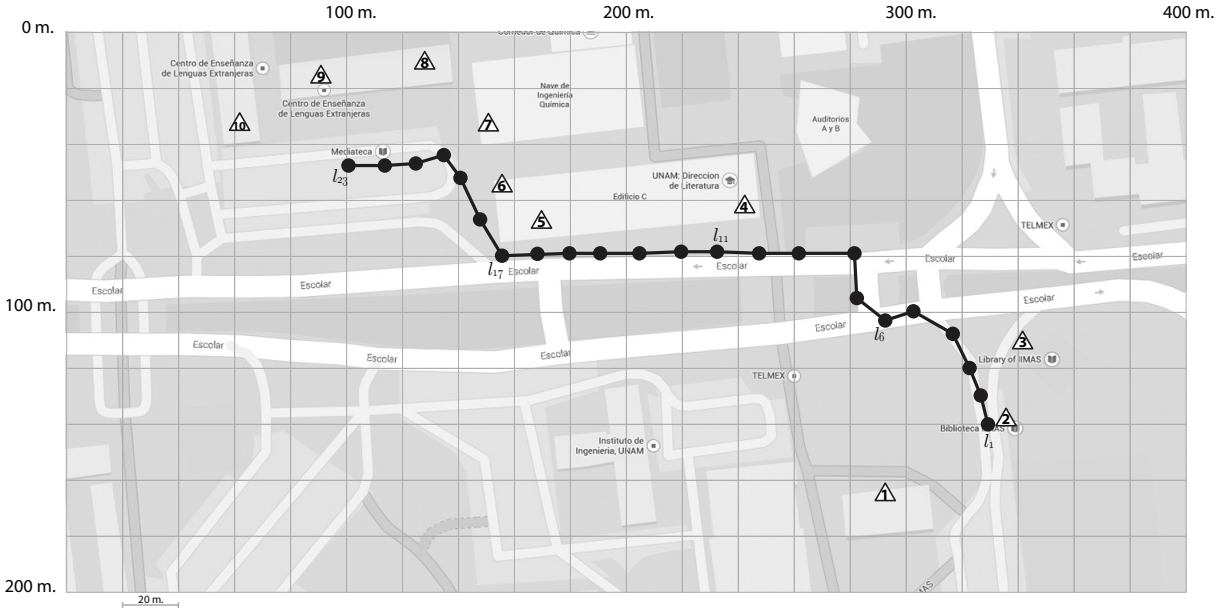


Figura 3.14: Prueba en el campus de Ciudad Universitaria. El origen del sistemas de coordenadas se ubica en la esquina superior izquierda (lat $19°19'54.2''$ N, long $99°11'2''$ W).

AP que escucharon la señal proveniente del nodo móvil. La tercera columna muestra la potencia de transmisión utilizada por el nodo móvil en el caso No-TPC. La cuarta y quinta columnas muestran el error de estimación de CPS y WCL, respectivamente. Asimismo, la sexta columna muestra el número promedio de AP que escucharon la señal proveniente del nodo móvil en el caso LTPC. La séptima columna muestra la potencia de transmisión utilizada por el nodo móvil en el caso LTPC. La octava y novena columnas muestran el error de estimación por CPS y WCL, respectivamente. Por ejemplo, para el caso No-TPC, en el renglón l_{21}, el número promedio de AP que escucharon al nodo móvil es 3.8, el error de estimación promedio por CPS es igual a $2.6\,\mathrm{m}$, mientras que el error de estimación promedio de WCL fue igual a $19.7\,\mathrm{m}$. En contraste, para la misma posición pero para el escenario LTPC, el número promedio de AP que escucharon al nodo móvil se redujo a 1.4, el error de estimación promedio en el caso CPS se incremento a $54.7\,\mathrm{m}$, mientras que el error de estimación promedio en el caso WCL se incrementó a $24.3\,\mathrm{m}$. También puede notarse, en el último renglón de esta tabla, que la diferencia del error de estimación promedio entre No-TPC y LTPC para CPS es igual a $38.12\,\mathrm{m}$, mientras que la diferencia de error de estimación promedio entre No-TPC y LTPC para WCL es igual a $5.11\,\mathrm{m}$. Este resultado indica que la eficacia de TPC está fuertemente relacionada con el algoritmo de localización utilizado por los atacantes. Además, solo en 3 de los 23 casos, el nodo móvil pudo ajustar su potencia de transmisión al valor óptimo requerido (ver séptima columna).

3.5. Resumen del capítulo

En este capítulo, primero se modeló el funcionamiento de la técnica de ofuscación TPC a través de un modelo probabilístico. Con este modelo se encontró una expresión analítica que le permite a un nodo móvil estimar la potencia de transmisión óptima de tal suerte que solo el AP más cercano escuche sus señales. Posteriormente, se analizó

cómo los algoritmos de localización CPS y WCL son afectados por TPC. Finalmente, se presentaron los resultados de las simulaciones y experimentos en condiciones realistas para evaluar la eficacia de TPC como técnica de privacidad geográfica. Estos resultados muestran que la eficacia de TPC depende principalmente de tres factores. Primero, la capacidad de un nodo móvil para ajustar su potencia de transmisión al valor óptimo. Segundo, las condiciones del canal inalámbrico como son: γ, σ y la densidad de AP. Tercero, el algoritmo de localización que los atacantes implementan. Sin embargo, aun cuando un nodo móvil pudiera superar el primer factor mediante dispositivos electrónicos más sofisticados, y el segundo mediante técnicas como diversidad en tiempo, espacio y frecuencia, es el tercer factor el que determina el verdadero valor de usar TPC como técnica de privacidad geográfica, debido a que el nodo móvil no tiene manera de saber qué algoritmo de localización se está usando para localizarlo. Por lo anterior podemos concluir que TPC no garantiza privacidad geográfica a los nodos móviles en entornos LE.

Capítulo 4

MSP

En este capítulo se presenta el protocolo denominado MAC *swapping protocol* (MSP), una estrategia que permite a dos nodos móviles intercambiar sus direcciones MAC de tal suerte que esto pase desapercibido para los atacantes que tienen acceso a la información de la capa física o a la capa MAC de los nodos involucrados. Para ello, MSP utiliza dos algoritmos complementarios denominados Safe-zone y VIME, los cuales operan en la capa física y capa MAC, respectivamente.

Aunque los atacantes en entornos LE pueden utilizar la información proveniente de capa 3 o superiores tales como: direcciones IP, información de los puertos o el estado de los sockets con el fin vincular la identidad del usuario con su posición, en esta tesis únicamente se estudiarán aquellos atacantes con acceso a la información de la capa física y la capa MAC, ya que consideramos que no es posible abarcar todas las capas del modelo OSI y sus contramedidas en un solo trabajo de investigación. Por esta razón, se asume que tanto los nodos móviles como los atacantes solo tienen acceso a la información proveniente de las capas uno y dos del modelo OSI.

4.1. Análisis de la capa física

En este trabajo, se asume que un grupo de atacantes continuamente analizan las transmisiones inalámbricas con el propósito de localizar e identificar a los nodos móviles. Estos atacantes tienen a su vez la capacidad de detectar comportamientos fuera de lo común que indiquen si el nodo móvil está intentando evadir el proceso de localización; por ejemplo, pueden percatarse de que un nodo móvil ha cambiado su dirección MAC (p. ej. si observan variaciones abruptas del parámetro RSSI entre dos paquetes consecutivos). Como contramedida a este escenario, en esta tesis proponemos el algoritmo Safe-zone. Este algoritmo permite que dos nodos móviles intercambien sus direcciones MAC sin dejar ningún indicio en capa física. El principio fundamental de este algoritmo es la coordinación de dos nodos móviles, previo y posterior al intercambio. No obstante, antes de continuar con la descripción del funcionamiento de Safe-zone, explicaremos la manera en la que los atacantes suelen detectar un intercambio de identidad en capa física.

4.1.1. Modelo del atacante

En este trabajo, se considera a un conjunto de AP que fungen como atacantes, los cuales están ubicados de manera aleatoria en la red. Estos AP tienen la capacidad de recolectar mediciones de RSSI provenientes de los nodos móviles que estén dentro de su área de cobertura. Estas mediciones son usadas para estimar la posición de los nodos móviles, así como para detectar cualquier variación fuera de lo común que indique un intercambio de direcciones MAC.

De acuerdo con [39], la potencia recibida (RSSI) decrece mientras la distancia que separa un transmisor de un receptor se incrementa. Esta relación generalmente se modela usando la ecuación 4.1 [49].

$$P_{Rx}(d)\ \mathrm{dBm} = P_{Tx} - PL(d_0) - 10\gamma log(\frac{d}{d_0}) + X_\sigma. \tag{4.1}$$

En donde P_{Tx} corresponde a la potencia de transmisión, P_{Rx} es la potencia recibida, $PL(d_0)$ es un valor de referencia que corresponde a la potencia recibida a la distancia d_0 (el valor más común para d_0 es $1\,\mathrm{m}$), d es la distancia euclidiana entre el transmisor y el receptor, γ es el exponente de pérdida de energía por la trayectoria, y finalmente X_σ es una variable aleatoria con media cero y desviación estándar igual a σ, esta variable representa el factor de desvanecimiento de la señal.

Por otro lado, la diferencia entre dos mediciones consecutivas de RSSI puede calcularse usando la ecuación 4.2, en donde el término P_{Rx_1} corresponde a la potencia recibida de un nodo móvil en el tiempo t ubicado a la distancia d_1, mientras que el término P_{Rx_2} corresponde a la potencia recibida del mismo nodo móvil en el tiempo $t + \Delta t$ a una distancia d_2

$$\Delta P_{Rx} = P_{Rx1} - P_{Rx2}. \tag{4.2}$$

Asumiendo que el valor de γ es el mismo para todos los nodos móviles dentro del área de cobertura del AP, y dado que $K = P_{Tx} - PL(d_0)$ es una constante, se puede sustituir la ecuación 4.1 en la ecuación 4.2, esto implica que ΔP_{Rx} se puede calcular como:

$$\Delta P_{Rx} = 10\gamma log(\frac{d_2}{d_1}) + X_\delta \tag{4.3}$$

en donde X_δ es una variable aleatoria con media cero y desviación estándar igual a $\delta = \sqrt{2}\sigma$, esta variable resulta de la diferencia entre dos variables aleatorias idénticamente distribuidas X_σ; ΔP_{Rx} es el valor absoluto de la diferencia de la potencia recibida desde dos posiciones distintas. Dado que el valor de γ es constante, esta ecuación depende únicamente del cociente ($\frac{d_2}{d_1}$). En consecuencia, un atacante que analice la información de capa física, esperaría que ΔP_{Rx} no tenga grandes variaciones entre dos mediciones consecutivas de RSSI provenientes de la misma dirección MAC. Esto debido a que se asume que todos los paquetes provienen del mismo nodo. La figura 4.1(a) muestra las mediciones de RSSI de dos nodos móviles que se desplazan dentro de la zona de cobertura 1-AP. En esta figura el nodo móvil *A* y el nodo móvil *B* intercambian sus direcciones MAC en el tiempo $t = 500\,\mathrm{s}$, de tal forma que después del intercambio, el nodo *A* transmite con la identidad del nodo *B* y viceversa. Desde el punto de vista del AP, el intercambio de direcciones MAC es detectado siempre y cuando el valor de ΔP_{Rx}, medido entre dos valores de RSSI consecutivos provenientes de la misma dirección MAC, exceda un umbral previamente definido [50].

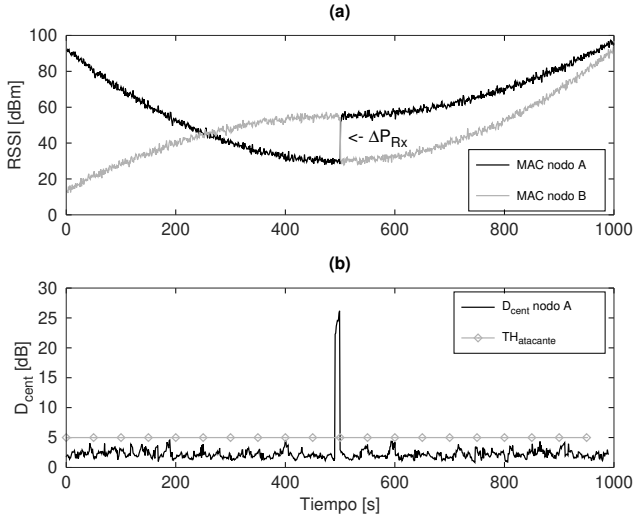


Figura 4.1: Funcionamiento del detector KSD.

Por otro lado, el detector de MAC *spoofing* propuesto en [36] utiliza el algoritmo k-means para determinar si los paquetes asociados a una misma dirección MAC provienen de una sola terminal inalámbrica o de varias. Es decir, este algoritmo forma $k \geq 2$ grupos a partir de mediciones RSSI provenientes de una misma dirección MAC de tal manera que si la distancia entre dos o más grupos es mayor a un cierto umbral, se considera como un evento de MAC *spoofing*. Una limitante de dicho algoritmo es que fue diseñado para nodos inalámbricos estáticos, sin embargo, este algoritmo puede modificarse para detectar variaciones abruptas del parámetro RSSI provenientes de nodos móviles. Esta modificación la denominamos KSD y consiste en utilizar una ventana deslizante de tamaño m a lo largo de mediciones consecutivas de RSSI provenientes de la misma dirección MAC. De esta manera, KSD considera un conjunto de mediciones de RSSI ($x_1, x_2, \ldots, x_m$). Posteriormente, KSD separa estos m valores de RSSI en k conjuntos, de tal suerte que las mediciones RSSI con las mismas características forman parte de un mismo grupo. Para detectar un intercambio de direcciones MAC, KSD debe estimar la distancia entre los centroides de cada grupo (D_{cent}). De esta manera, si dicha distancia excede un umbral predefinido ($\mathcal{TH}_{atacante}$), esto indica un intercambio de direcciones MAC. En la sección 4.4, se determina el valor de $\mathcal{TH}_{atacante}$ de manera empírica. La figura 4.1(b) muestra el funcionamiento de KSD. En esta figura se puede notar como el atacante puede detectar el momento en el cual el nodo A intercambia su identidad con el nodo B en el tiempo $500\,\mathrm{s}$ dado que D_{cent} excede el umbral $\mathcal{TH}_{atacante} = 5\,\mathrm{dB}$.

4.1.2. El algoritmo Safe-zone

En esta tesis se propone el algoritmo Safe-zone, un mecanismo que le permite a dos nodos móviles intercambiar su direcciones MAC sin dejar rastro para aquellos atacantes que analicen la información de la capa física.

El algoritmo Safe-zone funciona de la siguiente manera. El nodo móvil que desea intercambiar su dirección MAC envía una petición a sus nodos vecinos a través de un broadcast (MACRQ). Los nodos que reciban esta petición y deseen intercambiar su dirección MAC, deben responder mediante un paquete MACR. El nodo que inicia

el intercambio calcula $P_r\{\Delta P_{Rx} > \mathcal{TH}_{safe-zone}\}$ por cada MACR recibido, y selecciona aquellos candidatos que cumplan con $P_r \leq Pr_{max}$. En donde, P_r denota la probabilidad de que un atacante pueda detectar el intercambio, $\mathcal{TH}_{safe-zone}$ es la variación máxima de ΔP_{Rx} que puede pasar desapercibida por el atacante, y Pr_{max} es el límite que el nodo móvil fija con el fin de seleccionar al mejor candidato y, simultáneamente, descartar aquellos candidatos que pudieran revelar el intercambio. En este trabajo, se asume que el medio es simétrico (i.e. las mediciones RSSI recolectadas de un nodo móvil a un AP, son similares a las mediciones RSSI recolectadas del AP al nodo móvil).

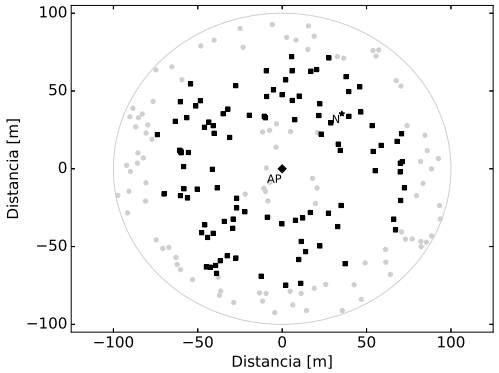
Las figuras 4.2(a), 4.2(b) y 4.2(c) muestran el funcionamiento de Safe-zone, para 1-AP, 2-AP y 3-AP, respectivamente. Los círculos de color negro representan el conjunto de posibles candidatos (i.e. los nodos con probabilidad por debajo de Pr_{max}). En estas figuras la posición del nodo móvil que desea intercambiar su dirección MAC se ejemplifica con una estrella, a partir de este momento este nodo será denominado iniciador. El nodo iniciador debe escoger dentro de este conjunto de candidatos al mejor de ellos negociando primero con aquel que tenga la P_r más baja. Si este candidato no acepta el intercambio, el iniciador consulta al siguiente candidato con la P_r más baja, y así sucesivamente. Si ninguno de los candidatos aceptan el intercambio de direcciones MAC, el iniciador debe comenzar todo este proceso pero en otra ubicación distinta.

Para calcular $P_r\{\Delta P_{Rx} > \mathcal{TH}_{safe-zone}\}$, el nodo iniciador debe estimar el valor de ΔP_{Rx} (ver ecuación 4.3) el cual depende de las distancias d_1 y d_2, en donde d_1 es la distancia entre el nodo iniciador y el AP, y d_2 es la distancia entre el nodo candidato y el AP. Para calcular d_1, el nodo iniciador recolecta un conjunto de muestras de RSSI provenientes de los paquetes enviados por el AP. De igual forma, los candidatos envían un conjunto de muestras de RSSI dentro del paquete MACR. Debido a que las distancias d_1 y d_2 son estimadas a partir de la ecuación 4.1, estas tienen variaciones por la variable aleatoria gaussiana. En consecuencia, estos vectores de distancias cuentan con un valor inferior y uno superior. Con estos valores Safe-zone debe calcular el peor caso que maximice la expresión $\left|10\gamma log(\frac{d_2}{d_1})\right|$, esto con el fin de garantizar que cualquier combinación entre d_1 y d_2 no exceda el límite impuesto por $\mathcal{TH}_{safe-zone}$. Para conseguir lo anterior, Safe-zone selecciona el valor más grande entre $\left|10\gamma log(\frac{d_{2_{upper}}}{d_{1_{lower}}})\right|$ y $\left|10\gamma log(\frac{d_{2_{lower}}}{d_{1_{upper}}})\right|$. Después, Safe-zone calcula $P_r\{\Delta P_{Rx} > \mathcal{TH}_{safe-zone}\}$ de la siguiente manera:

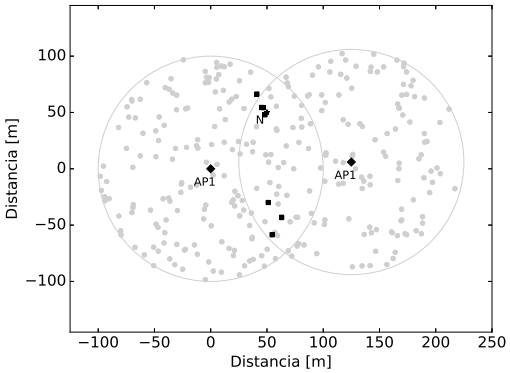
$$Pr\{\Delta P_{Rx} > \mathcal{TH}_{safe-zone}\} = \int_{\mathcal{TH}_{safe-zone}}^{\infty} \frac{1}{\sqrt{2\pi}\delta} e^{-\frac{(x-\mu)^2}{2\delta^2}} dx \tag{4.4}$$

en donde μ es el valor más grande de $\left|10\gamma log(\frac{d_2}{d_1})\right|$, δ es la desviación estándar cuyo valor es igual a $\sqrt{2}\sigma$, mientras que σ se obtiene de la ecuación 4.1. El algoritmo 1 muestra el funcionamiento de Safe-zone descrito en pseudo-código.

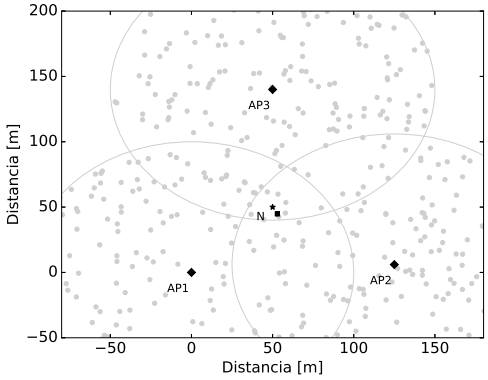
Finalmente, es importante mencionar que Safe-zone requiere de un canal privado para intercambiar todos sus paquetes de señalización (p. ej. MACR o MACRQ), sin que los atacantes sean capaces de escuchar la negociación por parte del iniciador y los candidatos. Para lograr esto, se pueden usar varios mecanismos. Por ejemplo, en [51], los autores proponen una metodología en la cual un usuario envía información oculta a los atacantes, usando los campos del encabezado de los paquetes. Un enfoque diferente es utilizar una modulación menos robusta (p. ej. 16QAM ó 64QAM), la cual dificulte la decodificación de los paquetes por parte del AP, esto debido principalmente



(a) Escenario 3-AP.



(b) Escenario 2-AP.



(c) Escenario 1-AP.

Figura 4.2: Candidatos potenciales determinados por Safe-zone para diferentes números de AP.

a la distancia que separa al nodo móvil del AP y al ruido del canal. Sin embargo, en esta tesis utilizaremos la técnica TPC en el escenario 1-AP [3] para enviar los paquetes de señalización de MSP. Para ello, se calcula TPC_p mediante $p(P_{Rx}(d_1) < P_{min})$. Esto con el fin de garantizar que la comunicación entre el iniciador y los candidatos no sea escuchada por el AP. Si la potencia óptima calculada por TCP_p está por debajo de la potencia mínima que el radio puede ajustar, entonces el nodo móvil tendrá que buscar otra posición donde pueda ajustar su potencia y hacer nuevamente el proceso para garantizar que el AP no sea capaz de decodificar sus paquetes. No obstante, el algoritmo Safe-zone puede funcionar con cualquier mecanismo que evite que los atacantes puedan escuchar los paquetes de señalización previos al intercambio de direcciones MAC.

4.2. Análisis de la capa MAC

Aun cuando dos nodos móviles puedan realizar un intercambio sin ser detectados en la capa física, en la capa MAC un atacante puede percatarse del intercambio,

Algoritmo 1 Safe-zone.

```
1: procedure MAIN LOOP
2:     while TRUE do
3:         Send MACRQ
4:         if candidate responds with MACR then
5:             Compute d_1
6:             for each candidate do
7:                 Compute d_2
8:                 Compute μ = max |ΔP_Rx|
9:                 Compute P_r{ΔP_Rx > TH_safe−zone}
10:                if P_r < Pr_max then
11:                    Append candidate to possible list
12:
13:        Sort possible candidates list by ascending order
14:        for each candidate in the list do
15:            Send MAC exchange Start-Negotiation
16:            if Candidate Accept then
17:                Send MAC exchange Confirm-Negotiation
18:                Update MAC address in VIME config
19:                Send MAC exchange End-Negotiation
20:                Break for loop
21:            else
22:                Send MAC exchange End-Negotiation
23:        Wait time interval t_0
```

ya que los sistemas operativos reinician la interfaz inalámbrica del nodo móvil cada vez que cambia su dirección MAC. Esto provoca que el número de secuencia del encabezado IEEE 802.11 tome el valor de cero cada que se modifica la dirección MAC. Los atacantes pueden detectar este comportamiento poco habitual y podrían advertir el intercambio de direcciones MAC al observar todos estos indicios que lo delatan. Como contramedida a esta amenaza, diseñamos VIME, un algoritmo que trabaja en la capa MAC y que hace uso de interfaces de red virtuales. VIME logra que los nodos móviles cambien sus direcciones MAC sin reiniciar los números de secuencia y sin reasociarse con el AP. No obstante, antes de explicar el funcionamiento de VIME, explicaremos cómo funciona, en general, un atacante en capa MAC.

4.2.1. Modelo del atacante

La mayoría de los detectores de robos de identidad (MAC *spoofing*) revisados en el capítulo 2 son incapaces de detectar el intercambio coordinado de direcciones MAC entre dos nodos móviles. Por ejemplo, en [33], los autores asumen que tanto el nodo legítimo como el que intenta robar la identidad transmiten de manera simultánea usando la misma dirección MAC. Este supuesto no afecta a MSP, ya que los nodos móviles se coordinan para no utilizar la misma dirección MAC simultáneamente. Sin embargo, los detectores de MAC *spoofing* propuestos en [38, 52] tienen mayor probabilidad de detectar el intercambio, dado que basan su funcionamiento en la observación de las discrepancias en el número de secuencia del encabezado IEEE 802.11. En específico, la estrategia SNG propuesta en [38] es la que tiene el mayor potencial de lograrlo, además de que es muy sencilla de implementar comparada con [52] ya que solo requiere calcular la diferencia en el número de secuencia entre dos paquetes consecutivos.

Un atacante en capa MAC puede verse como una función $y = f(SN_{gap})$, en donde SN_{gap} corresponde a la diferencia en el número de secuencia entre dos paquetes de datos consecutivos, mientras que $y \in (0, 1) \subset \mathbb{R}$. El valor de y expresa la cercanía de la diferencia SN_{gap} con respecto a un umbral dado, el cual denominaremos $SN_{threshold}$. Es decir, los valores de y cercanos a cero indican que el valor de SN_{gap} es menor a $SN_{threshold}$, mientras que valores cercanos a uno indican que SN_{gap} es mayor a $SN_{threshold}$. La función y que nosotros vamos a utilizar en esta tesis se muestra en 4.5, sin embargo cualquier otra función que cumpla estos requerimientos puede ser utilizada.

$$y = f(SN_{gap}) = 1 - e^{-\frac{ln(2)SN_{gap}^2}{SN_{threshold}^2}} - e^{-\frac{ln(2)(SN_{gap}-4096)^2}{SN_{threshold}^2}}. \tag{4.5}$$

El valor del umbral $SN_{threshold}$ se determinó experimentalmente. En dichos experimentos se observó que en el $99.97\,\%$ de los casos el valor SN_{gap} se encuentra entre 1 y 25. Por lo tanto si fijamos el valor de $SN_{threshold}$ igual a 25, el detector presetenaría sólo un $3\,\%$ de falsos positivos. Cabe resaltar que el valor de SN_{gap} suele ser mayor a uno debido a las colisiones entre paquetes de datos en el canal inalámbrico. En consecuencia, es posible construir un detector de capa MAC a partir de la función 4.5. Por ejemplo, si consideramos una alarma cada vez que el valor de y exceda el valor de 0.5, esto implica que el valor SN_{gap} ha excedido $SN_{threshold}$. Contrario a la referencia [38], nuestro detector también considera el caso cuando el número de secuencia llega a su valor límite (i.e. 4096) y vuelve a tomar el valor de cero, el cual es un comportamiento normal [37]. Este comportamiento es contemplado en el segundo

Tabla 4.1: Valores de los campos tipo y subtipo en el encabezado IEEE 802.11.

Tipo b3 b2	**Descripción**	**Subtipo b7 b6 b5 b4**	**Descripción**
00	Control	0000	*Association request*
00	Control	0001	*Association response*
00	Control	0010	*Reassociation request*
00	Control	0011	*Reassociation response*
00	Control	0100	*Probe request*
00	Control	0101	*Probe response*
00	Control	1010	*Disassociation*
00	Control	1011	*Authentication*
00	Control	1100	*Deauthentication*

término de la ecuación 4.5. Además, le permite a nuestro detector reducir el número de falsos positivos mientras se analiza el incremento de los números de secuencia. Este detector se muestra en la ecuación 4.5.

Ahora abordaremos la detección del intercambio de direcciones MAC debido al proceso de reasociación. De acuerdo con el estándar IEEE 802.11, cada vez que un nodo móvil se asocia con un AP se transmiten un conjunto de paquetes de control. Por esta razón, mejoramos el modelo de atacante al incluir un detector de eventos de reasociación. En específico, en la tabla 4.1 se muestran los campos *tipo* y *subtipo* del encabezado IEEE 802.11, así como los valores usados en los paquetes del proceso de asociación. Con base en estos campos, modelamos nuestro detector como una función que considera los valores del campo tipo y subtipo, esta función debe comparar si las combinaciones de estos dos campos aparecen en el encabezado IEEE 802.11 de cualquier paquete transmitido; si la combinación indica un paquete de control relativo a la asociación, la salida de nuestra función será un uno, de lo contrario, la salida de la función será un cero.

4.2.2. El algoritmo VIME

Debido a que los atacantes pueden detectar un evento de MAC *spoofing* mediante la utilización de la técnica SNG, en esta tesis proponemos el algoritmo VIME como una contramedida que permite a los usuarios móviles modificar cualquier campo en el encabezado IEEE 802.11 antes de entregar los paquetes de datos a la interfaz inalámbrica. Para lograr esto, VIME utiliza interfaces virtuales de red. Gracias a estas interfaces, VIME puede cambiar cualquier campo en el encabezado IEEE 802.11 sin reiniciar la interfaz inalámbrica evitando de esta forma cualquier comportamiento fuera de lo normal durante la comunicación con el AP.

En sistemas operativos basados en UNIX, los usuarios disponen de un driver llamado TUN/TAP [53]. Este driver le permite a un usuario redireccionar los paquetes provenientes del sistema operativo hacia un descriptor de archivo (ubicado en

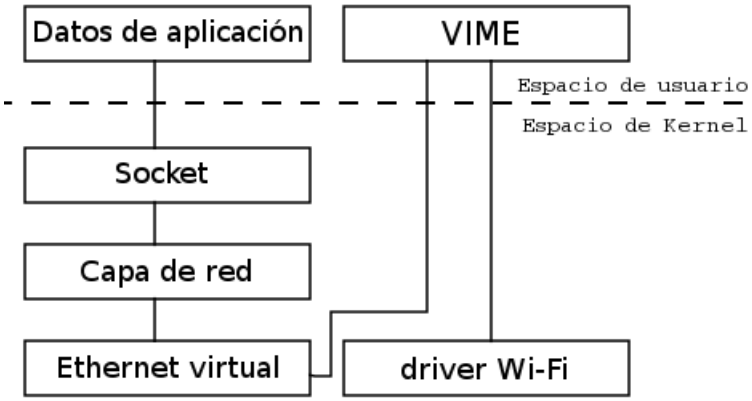


Figura 4.3: Relación entre VIME y los componentes de red de un sistema Unix.

/dev/net/tun) en vez de enviarlos a la interfaz física de red. Este driver puede configurarse de dos maneras. La primera se conoce como configuración TUN. En esta, el driver crea un túnel a través de una interfaz lógica (interfaz virtual), la cual envía el tráfico entre dos puntos de red. Por otro lado, en su configuración TAP, el driver crea una interfaz Ethernet virtual la cual recibe los paquetes provenientes de las aplicaciones del usuario encapsulados en el formato IEEE 802.3. Estos paquetes se escriben en el descriptor de archivo. VIME aprovecha la configuración TAP para desarrollar una aplicación de usuario que pueda modificar fácilmente el encabezado de un paquete antes de entregarlo a la NIC inalámbrica. De esta manera, VIME se convierte en un intermediario entre la interfaz Ethernet virtual y la tarjeta de red inalámbrica.

La figura 4.3 muestra un diagrama de la relación entre VIME y los componentes principales de red de un sistema Unix. En este diagrama, el kernel de Unix crea un socket que sirve como interfaz entre las aplicaciones del usuario y los protocolos de red. El bloque Capa de red implementa todo lo relacionado con el modelo de comunicación TCP/IP. En particular, este bloque se encarga de ejecutar las funciones necesarias para encapsular los paquetes de datos usando los protocolos TCP/IP, de igual manera, este bloque se encarga de las funciones de encaminamiento. Finalmente, el kernel envía los paquetes IP a la interfaz Ethernet virtual. En donde VIME los puede modificar a través del descriptor de archivo ubicado en “/dev/net/tun" con el fin de cambiar el encabezado IEEE 802.3 del paquete original por uno del tipo IEEE 802.11. Es en este paso donde VIME tiene la posibilidad de modificar cada uno de los campos del encabezado antes de enviarlo a la interfaz inalámbrica. En específico, VIME modifica los campos dirección MAC de origen y número de secuencia para que coincidan con aquellos valores que recibió del candidato seleccionado por Safe-zone al momento de llevar a cabo el intercambio de direcciones MAC. Esto con el fin de evitar cualquier detección por parte de los atacantes que están observando la información de capa MAC. Una vez concluye el proceso de encapsulado, VIME inyecta cada paquete modificado directamente a una interfaz inalámbrica para su transmisión. Este proceso de actualizar los valores de dirección MAC y número de secuencia se lleva a cabo de manera instantánea, por lo que no hay pérdida alguna de paquetes salientes durante el proceso.

Es importante mencionar que VIME funciona en ambos sentidos, es decir, cuando se crea un paquete de la capa de aplicación y se envía a la interfaz Wi-Fi (ver figura

4.3), así como cuando el paquete entra por la interfaz Wi-Fi y sube a la capa de aplicación. En este último caso, VIME prepara los paquetes recibidos antes de enviarlos a la interfaz Ethernet virtual reemplazando el encabezado IEEE 802.11 con el encabezado IEEE 802.3. Una ventaja de usar interfaces virtuales es que VIME se ejecuta como una aplicación de usuario y no requiere modificación alguna del sistema de red de los sistemas Unix. Se muestra el pseudocódigo de VIME en el algoritmo 2.

Algoritmo 2 VIME.

```
 1: procedure MAIN LOOP
 2:     Point to the file descriptor /dev/net/tun
 3:     Create raw socket on wireless interface
 4:     while TRUE do
 5:         if packet in /dev/net/tun then
 6:             Read packet
 7:             Remove Ethernet header
 8:             Append Wi-Fi header
 9:             Send packet to wireless interface
10:         if packet in wireless interface then
11:             Read packet
12:             if packet is for this terminal then
13:                 Remove Wi-Fi header
14:                 Append Ethernet header
15:                 Send packet to /dev/net/tun
```

4.3. Intercambio de identidad en MSP

Aun cuando un solo intercambio de direcciones MAC puede confundir a los atacantes, estos pueden reunir suficiente información a lo largo del tiempo para identificar la identidad real del usuario por otros medios. Por ello, es recomendable que los nodos móviles intercambien periódicamente sus direcciones MAC (i.e. identidad). La pregunta inmediata es cuánto tiempo debe durar un identificador. La respuesta a esta pregunta sigue siendo un tema de investigación que aún no tiene respuesta definitiva [54, 55]. Sin embargo, existen varias técnicas que han sido propuestas en la literatura para tratar de resolver este problema. Por ejemplo, en la referencia [32], los autores modelaron el tiempo que debe durar un seudónimo como la probabilidad de encontrar al menos un candidato para el intercambio, así como el costo del intercambio. En la referencia [55], los autores proponen que la frecuencia de intercambio de seudónimos en redes vehiculares sea inversamente proporcional al intervalo de tiempo en el cual el vehículo es observado por los atacantes. Otra estrategía podría considerar el número de AP escuchando la señal del nodo móvil como factor para establecer la duración del seudónimo. En general, entre más atacantes estén escuchando la señal del nodo móvil, más exacta será su localización, y por consiguiente aumenta la necesidad del usuario de intercambiar su identidad. En la referencia [56], los autores proponen un mecanismo que le permite a un nodo móvil medir con qué exactitud los atacantes estiman su posición. MSP puede utilizar este enfoque para ajustar la duración de los

Tabla 4.2: Términos utilizados en Safe-zone y VIME.

Parámetro	Descripción
D_{cent}	Distancia entre centroides, utilizado en el algoritmo KSD.
$\mathcal{TH}_{attacker}$	Umbral del detector de capa física.
$\mathcal{TH}_{safe-zone}$	Valor máximo permitido de la variación ΔP_{Rx} considerada por el algoritmo Safe-zone.
P_r	Probabilidad de ser detectado por los atacantes cuando se lleva a cabo un intercambio de direcciones MAC.
Pr_{max}	Probabilidad máxima permitida para pasar desapercibido por los atacantes.
SN_{gap}	Salto en el número de secuencia entre dos paquetes consecutivos.
$SN_{threshold}$	Valor del umbral considerado por el atacante SNG en capa MAC.

seudónimos de acuerdo al número de atacantes que tenga a su alrededor el nodo móvil. No obstante, MSP puede utilizar cualquier algoritmo que le indique a Safe-zone cada cuando buscar candidatos para el intercambio.

Finalmente, cabe resaltar que el nodo móvil que inicia el intercambio solo enviará paquetes MACRQ cuando el tiempo de vida de su seudónimo haya terminado o esté a punto de caducar. Asimismo, los candidatos solo responderán con el paquete MACR cuando el tiempo de vida de su seudónimo también haya expirado. Esto garantiza que ningún nodo efectúe múltiples intercambios de direcciones MAC, aun cuando reciba múltiples peticiones MACRQ por parte de varios iniciadores.

4.4. Pruebas y experimentos

En esta sección se describe la metodología utilizada para implementar tanto a los atacantes como las contramedidas propuestas en este capítulo. Posteriormente, se presentan los experimentos realizados tanto en la capa física, así como en la capa MAC con el objetivo de evaluar la eficacia de MSP para proveer privacidad geográfica a los nodos móviles en entornos LE. Finalmente, se presenta un análisis de seguridad y desempeño de MSP, así como una comparación de MSP con modelos de atacantes y contramedidas similares.

La tabla 4.2 muestra una breve descripción de los términos utilizados para el algoritmo Safe-zone, así como para el algoritmo VIME. El lector puede encontrar una descripción detallada en las secciones 4.1.2 y 4.2.2.

Como se describió previamente, el algoritmo Safe-zone debe calcular la probabi-

Tabla 4.3: Valor del umbral $\mathcal{TH}_{attacker}$ vs. el porcentaje de falsos positivos.

$\mathcal{TH}_{attacker}$	**Porcentaje de falsos positivos**	
	AP	Nodo A
1	0.64	0.68
2	0.32	0.35
3	0.14	0.18
4	0.05	0.09
5	0.03	0.05
6	0.01	0.03
7	0.009	0.02
8	0.004	0.014
9	0.0023	0.012
10	0.0021	0.010
15	0.0004	0.005

lidad $P_r\{\Delta P_{Rx} > \mathcal{TH}_{safe-zone}\}$ para seleccionar no solo la mejor ubicación, sino también al mejor candidato para llevar a cabo un intercambio de direcciones MAC. Sin embargo, para calcular $P_r\{\Delta P_{Rx} > \mathcal{TH}_{safe-zone}\}$ se requiere conocer el valor de los parámetros γ, K y σ para estimar el valor de ΔP_{Rx}. Los valores de estos parámetros se obtuvieron experimentalmente, a través de un AP ubicado en un ambiente exterior. El experimento consistió en un nodo móvil recolectando muestras de RSSI desde diferentes distancias respecto al AP. El nodo móvil utilizó un radio Wi-Fi con chip Atheros basado en el estándar IEEE 802.11n en la banda de los $2.4\,\mathrm{GHz}$. La captura de paquetes se llevó a cabo mediante el software *tcpdump*. Las distancias consideradas estuvieron dentro del rango de 5 a $100\,\mathrm{m}$ con incrementos de $5\,\mathrm{m}$. Se utilizó el modelo propagación path-loss mostrado en la ecuación 4.1 sobre los datos experimentales [57], con lo cual se obtuvieron los siguientes valores de los parámetros: γ igual a 1.34, K igual a $-44.12\,\mathrm{dB}$ y σ igual a $2.23\,\mathrm{dBm}$.

Con la finalidad de evaluar la eficacia de nuestro modelo de atacante, en este trabajo definimos la métrica falsos positivos para el atacante como el porcentaje de experimentos en los cuales el atacante detectó un intercambio de direcciones MAC en donde no lo hubo. El atacante usará esta métrica para seleccionar el mejor umbral $\mathcal{TH}_{atacante}$ que minimiza el porcentaje de falsos positivos. Para calcular $\mathcal{TH}_{atacante}$, realizamos dos experimentos.

En el primer experimento, caracterizamos la variación de D_{cent}, la cual se obtuvo al ejecutar el algoritmo KSD sobre un conjunto de mediciones RSSI obtenidas de un nodo móvil A que no intercambia su dirección MAC. Esto se hace con el fin de caracterizar el comportamiento típico de RSSI sin intercambios de identidad. Para este experimento, usamos un AP y un nodo móvil que se desplaza dentro del área de cobertura del AP (el alcance de comunicación máximo fue de aproximadamente $100\,\mathrm{m}$). El nodo móvil se desplazó siguiendo trayectorias rectilíneas a diferentes distancias del AP. El nodo móvil generó tráfico usando la herramienta *ping*, la cual fue configurada

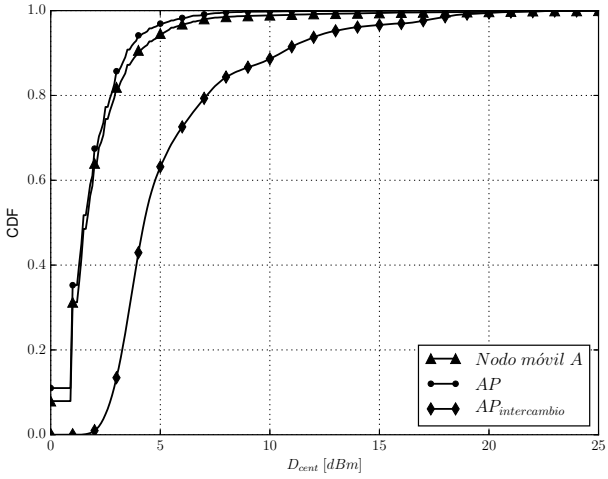


Figura 4.4: CDF de D_{cent}.

para enviar un paquete al AP cada $100\,\mathrm{ms}$. Dado que el protocolo ICMP incluye los paquetes *echo request* y *echo response*, cada paquete constituye una muestra tanto para el AP como para el nodo móvil. El AP recolectó 1×10^5 muestras de RSSI provenientes de paquetes transmitidos por el nodo móvil. Fijamos el tamaño de la ventana $m = 10$ y establecimos el valor de $k = 2$ para el algoritmo KSD[1]. La figura 4.4 muestra la CDF de D_{cent} para las medidas tomadas por el AP (línea con círculos). De manera similar, el nodo móvil también recolectó el mismo número de muestras RSSI provenientes del AP para caracterizar ΔP_{Rx} mediante el algoritmo KSD. En esta misma figura también se muestra la CDF de D_{cent} para las medidas tomadas por el nodo móvil (línea con triángulos). Como se puede observar en la misma figura, la CDF obtenida por el AP y la CDF obtenida por el nodo móvil son similares, lo que sustenta nuestra aseveración de que, en términos del promedio de mediciones, el canal inalámbrico puede considerarse simétrico.

En el segundo experimento, consideramos a un segundo nodo móvil (nodo *B*) el cual también se desplaza dentro del área de cobertura del AP siguiendo una trayectoria rectilínea. Dichas trayectorias se cruzan con las del nodo *A* en un sólo punto. Del mismo modo que en el experimento anterior, se recolectaron 1×10^2 muestras de mediciones de RSSI mientras el nodo *A* intercambia su dirección MAC con el nodo *B*. Al mismo tiempo el AP calcula la D_{cent} para cada nodo móvil. La figura 4.4 muestra la CDF de D_{cent} medida por el AP cuando los nodos móviles llevan a cabo el intercambio de direcciones MAC. En esta figura se puede ver que si el atacante selecciona el umbral $\mathcal{TH}_{atacante} = 15\,\mathrm{dB}$, solo el $4\,\%$ de los intercambios de dirección MAC son detectados (línea con diamantes). Al mismo tiempo, el AP tiene un falso positivo del $0\,\%$, esto se puede observar en la línea con círculos, en la cual todos los valores D_{cent} obtenidos sin intercambios de direcciones MAC se encuentran por debajo de dicha línea. Más aún, en la figura podemos observar que si el AP escoge $\mathcal{TH}_{atacante} = 1\,\mathrm{dB}$, esto implica que el AP puede detectar un intercambio de direcciones MAC con el $100\,\%$ de eficacia (línea con diamantes), pero al mismo tiempo el AP tendría un $64\,\%$ de falsos positivos (línea con círculos). Esto significa que el $64\,\%$ de todos los eventos

[1]El número de clusters.

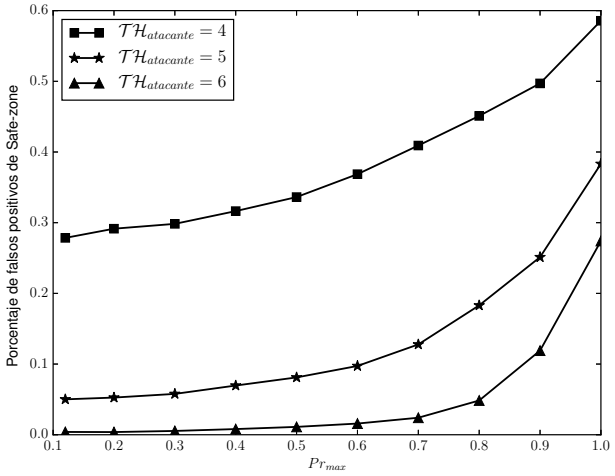


Figura 4.5: Porcentaje de falsos positivos de Safe-zone vs. Pr_{max}.

registrados por el AP no pueden diferenciarse entre los verdaderos intercambios de direcciones MAC y los falsos positivos. La tabla 4.3 muestra diferentes valores del umbral $\mathcal{TH}_{atacante}$ así como el porcentaje de falsos positivos correspondiente. A través de los resultados mostrados, consideramos que el umbral $\mathcal{TH}_{atacante} = 5\,\mathrm{dB}$ minimiza el porcentaje de falsos positivos y al mismo tiempo maximiza el número de detecciones de intercambios de direcciones MAC, ya que para este valor de umbral, el atacante puede detectar el $37\,\%$ de los intercambios reales de direcciones MAC con solo un $3\,\%$ de falsos positivos. De igual manera, un nodo móvil que use el algoritmo Safe-zone puede seleccionar su umbral ($\mathcal{TH}_{safe-zone}$) basado en la tabla 4.3.

Con la finalidad de evaluar la eficacia del algoritmo Safe-zone, en este capítulo definimos el porcentaje de falsos positivos de Safe-zone como el número de casos en los que el atacante detectó el intercambio aun cuando Safe-zone escogió al mejor candidato. Para cuantificar el porcentaje de falsos positivos, utilizamos el segundo experimento previamente descrito. En este experimento el nodo *A* es el iniciador del intercambio de direcciones MAC y el nodo *B* es el único candidato. Durante el intercambio de direcciones MAC, el nodo iniciador calcula $P_r\{\Delta P_{Rx} > \mathcal{TH}_{safe-zone}\}$, mientras que el atacante calcula la D_{cent}. Para este experimento realizamos 1×10^2 intercambios de direcciones MAC, utilizando el umbral $\mathcal{TH}_{safe-zone}$ igual a $5\,\mathrm{dB}$. La figura 4.5 muestra la variación del porcentaje de falsos positivos de Safe-zone con respecto a Pr_{max} y a $\mathcal{TH}_{atacante}$. En esta figura, se puede ver que entre más grande el valor de Pr_{max}, más grande el porcentaje de falsos positivos de Safe-zone. Por ejemplo, supongamos que el atacante escoge el valor del umbral $\mathcal{TH}_{atacante} = 4\,\mathrm{dB}$ mientras que el algoritmo Safe-zone escoge el valor de $Pr_{max} = 0.2$, en consecuencia el atacante podría detectar el $29.14\,\%$ de los intercambios de direcciones MAC (ver figura 4.5). Por otro lado, supongamos que el atacante y el iniciador escogen el mismo valor de umbral ($\mathcal{TH}_{safe-zone} = \mathcal{TH}_{atacante} = 5\,\mathrm{dB}$), La figura 4.5 muestra que cuando $Pr_{max} = 0.2$, solamente el $5.2\,\%$ de los intercambios de direcciones MAC serían detectados. También se puede ver en esta figura que el porcentaje de falsos positivos decrece cuando $\mathcal{TH}_{atacante} > \mathcal{TH}_{safe-zone}$.

La figura 4.6 muestra la CDF de los valores recolectados de P_r durante el segundo experimento. Aquí se puede observar que el $29.43\,\%$ de los casos quedaron por debajo

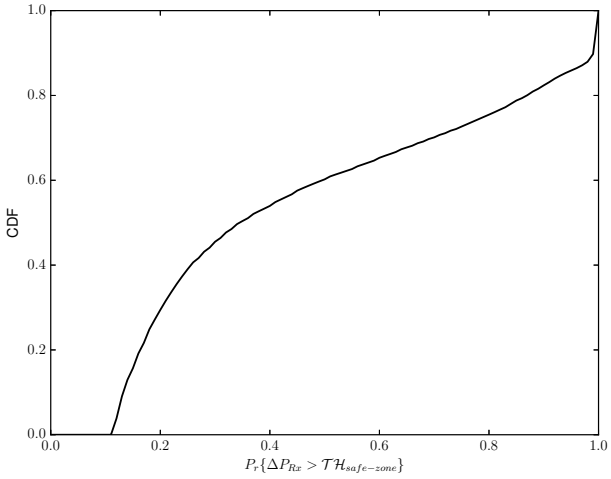


Figura 4.6: CDF de los valores P_r obtenidos en el experimento 2.

de $Pr_{max} = 0.2$. En otras palabras, un nodo móvil que se desplaza dentro del área de cobertura de un AP tiene una probabilidad de encontrar a un candidato con $P_r \leq 0.2$ es casi el $30\,\%$. En contraste, al considerar una $Pr_{max} = 0.4$, el nodo móvil tiene un $53.95\,\%$ de probabilidad de encontrar a un candidato. Sin embargo, la figura 4.5 muestra el hecho de que si Safe-zone selecciona $Pr_{max} = 0.4$ y $\mathcal{TH}_{safe-zone} = 5\,\mathrm{dB}$, y al mismo tiempo el atacante selecciona el valor de $\mathcal{TH}_{attacker} = 4\,\mathrm{dB}$, la eficacia de Safe-zone sería de $68.37\,\%$. Para los experimentos subsecuentes, utilizaremos el valor $Pr_{max} = 0.2$, dado que de acuerdo a los resultados mostrados en las figuras 4.5 y 4.6, Safe-zone tiene un $5\,\%$ de falsos positivos mientras que tiene un $30\,\%$ de probabilidad de encontrar un candidato idóneo para realizar el intercambio.

Con base en los resultados obtenidos en el segundo experimento, en la figura 4.7(a) se muestra la P_r calculada sobre todos los candidatos posibles, cuando el nodo iniciador del intercambio se encuentra a $15\,\mathrm{m}$ de separación con respecto al AP. En esta figura puede verse que, de acuerdo a Safe-zone, los mejores candidatos son los que están localizados entre 14 a $19\,\mathrm{m}$ y de 22 a $26\,\mathrm{m}$ respecto al AP. Esto debido a que los valores de P_r son menores que $Pr_{max} = 0.2$, cuando $\mathcal{TH}_{safe-zone} = 5\,\mathrm{dB}$. Más aún, si el atacante escogiera el umbral $\mathcal{TH}_{atacante} = 5\,\mathrm{dB}$, los nodos ubicados entre las distancias de 8 a $10\,\mathrm{m}$ y de 15 a $28\,\mathrm{m}$ podrían realizar intercambios de direcciones MAC sin ser detectados por los atacantes (ver figura 4.7(b)). Asimismo, la figura 4.8(a) muestra la probabilidad (P_r) cuando el nodo iniciador se encuentra ubicado a $50\,\mathrm{m}$ del AP. Como puede verse en esta figura, de acuerdo con Safe-zone los mejores candidatos ahora se encuentran ubicados entre 35 y $75\,\mathrm{m}$ con respecto al AP, dado que en esta zona P_r es menor que $Pr_{max} = 0.2$, cuando $\mathcal{TH}_{safe-zone} = 5\,\mathrm{dB}$. Por otro lado, si el atacante escogiera $\mathcal{TH}_{atacante} = 5\,\mathrm{dB}$, todos los nodos candidatos seleccionados por Safe-zone pasarían inadvertidos por los atacantes dado que la distancia D_{cent} es menor que $\mathcal{TH}_{atacante}$ (ver figura 4.8(b)). En contraste, si el atacante escogiera el umbral $\mathcal{TH}_{atacante} = 2\,\mathrm{dB}$, ninguno de los intercambios de direcciones MAC podría pasar desapercibido; sin embargo, el atacante observaría $32\,\%$ de todos los eventos como falsos positivos (ver tabla 4.3).

Finalmente, combinamos los algoritmos Safe-zone y VIME en un escenario de prueba para evaluar la eficacia de MSP como una técnica de privacidad geográfica.

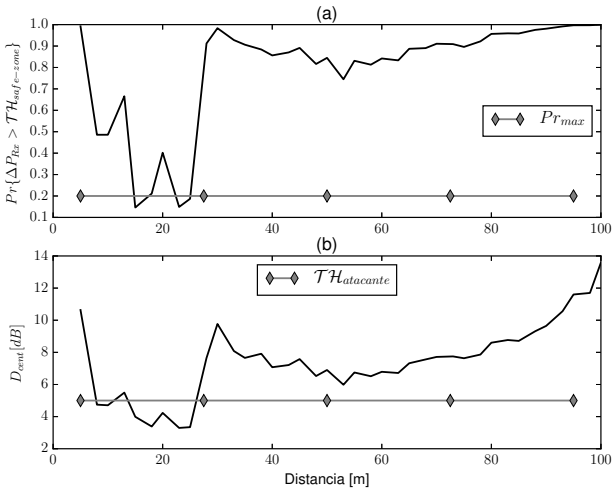


Figura 4.7: Candidatos potenciales de Safe-zone cuando el iniciador se encuentra a 15 m con respecto al AP.

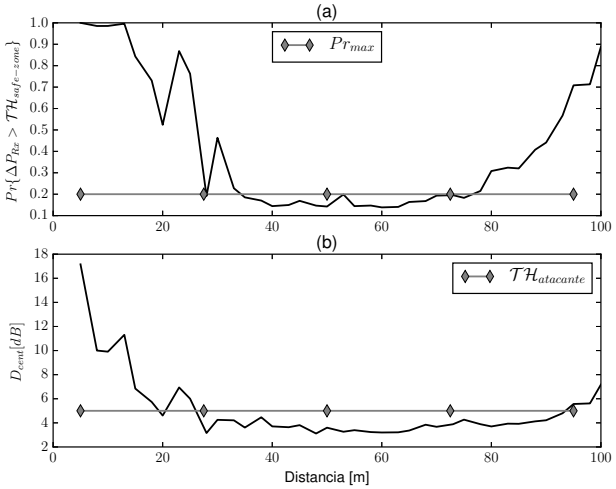


Figura 4.8: Candidatos potenciales de Safe-zone cuando el iniciador se encuentra a 50 m con respecto al AP.

En este escenario de prueba, dos nodos móviles intercambian sus direcciones MAC con y sin hacer uso de MSP. La figura 4.9(a) muestra las mediciones RSSI recolectadas por el AP provenientes de dos nodos móviles desplazándose dentro de su área de cobertura. En el tiempo $t = 2208\,\mathrm{s}$, ambos nodos llevan a cabo el intercambio de direcciones MAC sin hacer uso de MSP. En la figura 4.9(b), se puede observar cómo los valores del número de secuencia provenientes de ambos nodos móviles regresan a cero al momento del intercambio. Más aún, en la figura 4.9(c) puede verse cómo el detector de capa física evidencia el intercambio dado que el valor de D_{cent} supera $\mathcal{TH}_{atacante} = 5\,\mathrm{dB}$, para ambos nodos móviles. La figura 4.9(d) muestra cómo el salto en el número de secuencia supera el umbral de 25 de acuerdo con la ecuación 4.5. Finalmente, la figura 4.9(e) muestra la presencia de paquetes de control debido al

proceso de reasociación con el AP después de que los nodos móviles actualizaron su dirección MAC. Este escenario de prueba ejemplifica un intercambio de direcciones MAC que activa todos los detectores que el atacante puede considerar dado que los nodos móviles no utilizaron ninguna contramedida.

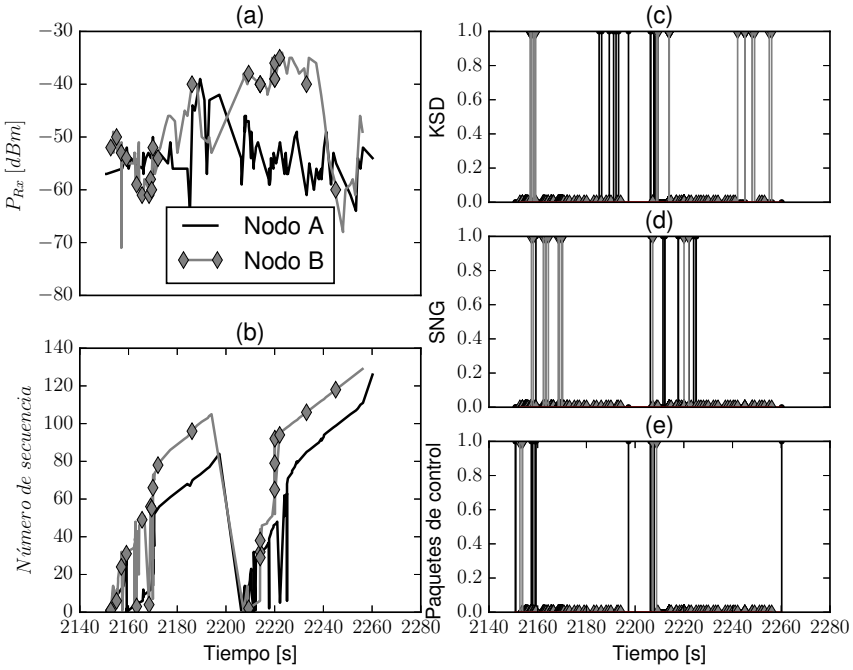


Figura 4.9: Intercambio de direcciones MAC sin hacer uso de MSP.

Con la finalidad de reducir el número de falsos positivos, los atacantes podrían combinar los detectores de capa física con los detectores de capa MAC como se muestra en la tabla 4.4. Por ejemplo, cuando el detector de saltos de número de secuencia y el detector de paquetes de control toman el valor de uno y al mismo tiempo, el detector de capa física toma el valor de cero, indicaría que hubo un reinicio de la interfaz inalámbrica del nodo móvil. La combinación de estos tres detectores disminuye el porcentaje de falsos positivos del atacante, lo que incrementa la probabilidad de detectar el intercambio real de direcciones MAC.

La figura 4.10 muestra el escenario cuando dos nodos móviles intercambian sus direcciones MAC utilizando MSP. En esta figura puede verse que los valores del número de secuencia asociado a la dirección MAC de cada nodo móvil muestran una tendencia creciente, sin activar el detector. Aun cuando el intercambio de direcciones MAC se realizó en el tiempo $t = 1448\,\mathrm{s}$, el detector de números de secuencia mantuvo su valor en cero. Para este experimento, el nodo iniciador se ubicó a $30\,\mathrm{m}$ con respecto al AP, mientras que el candidato se ubicó a $50\,\mathrm{m}$ con respecto al AP. Los parámetros utilizados por el iniciador fueron $Pr_{max} = 0.2$ y $\mathcal{TH}_{safe-zone} = 5\,\mathrm{dB}$. Este escenario ejemplifica un intercambio de direcciones MAC que no deja ningún indicio aun cuando el detector de capa física y los de capa MAC trabajaron simultáneamente.

Tabla 4.4: Combinación de los detectores de capa física y capa MAC.

KSD	SNG	Paquetes de control	Descripción
0	0	0	Operación normal.
0	0	1	Comportamiento fuera de lo normal.
0	1	0	Pérdida de paquetes o MAC *spoofing*.
0	1	1	Reinicio de la interfaz inalámbrica o posiblemente un intercambio de identidad.
1	0	0	Variciones del parámetro RSSI o MAC *spoofing*.
1	0	1	Comportamiento fuera de lo normal.
1	1	0	MAC *spoofing*.
1	1	1	Intercambio de direcciones MAC.

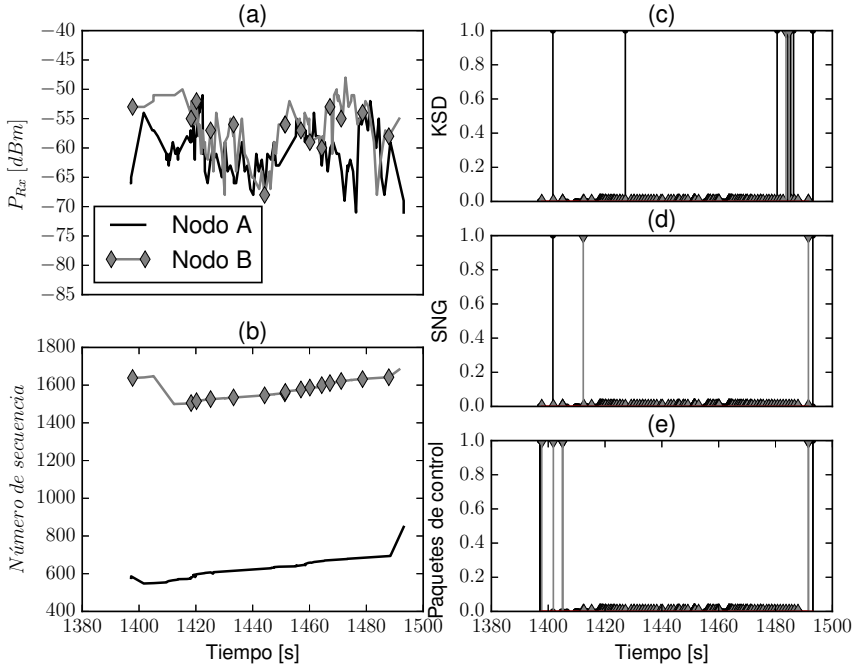


Figura 4.10: Intercambio de direcciones MAC haciendo uso de MSP.

Tabla 4.5: MSP vs atacantes.

Capa	Atacante	Protección de MSP
Física	Trilateración [58]	Dado que MSP puede disociar la ubicación del usuario de su identidad, los algoritmos de trilateración obtendrían una posición errónea (i.e. de otro usuario).
Física	MAC *spoofing* [34, 35, 36, 33]	Dado que el algoritmo Safe-zone selecciona al mejor candidato mediante el análisis de los valores RSSI, desde la persepectiva del atacante, elimina todo comportamiento fuera de lo normal en la capa física (i.e., saltos en los valores de RSSI).
MAC	MAC *spoofing* [38, 33]	Dado que el algoritmo VIME puede modificar los encabezados de la capa MAC (dirección MAC de origen y número de secuencia) entre el nodo candidato y el iniciador, el intercambio de identidad no deja ningún indicio ni muestra signos de comportamiento fuera de lo normal (i.e. saltos en el número de secuencia).

4.4.1. Análisis de seguridad

Esta sección presenta una comparación de MSP contra los atacantes existentes en la literatura, en específico, aquellos que utilizan la información de la capa física o de la capa MAC. La tabla 4.5 muestra estos atacantes, así como una discusión explicando si MSP es o no capaz de engañarlos. Como puede verse en esta tabla, ninguno de los atacantes reportados fue capaz de detectar un intercambio de direcciones cuando los nodos móviles utilizaron MSP. Incluso, cuando se considera el caso en el que el AP (i.e. atacante) pretende ser un candidato o el iniciador de un intercambio de direcciones MAC con fin de invalidar la estrategia. En estos casos, el algoritmo Safe-zone descartaría a este candidato debido a que nodos ubicados cerca del AP tienen una P_r por encima del umbral $\mathcal{TH}_{safe-zone}$. Esto se puede observar en las figuras 4.7(a) y 4.8(a) en las cuales el valor de P_r es cercano a uno para aquellos candidatos ubicados más cerca del AP.

Por otro lado, si consideramos el escenario en el que un nodo móvil es un atacante. MSP no puede distinguir si algún nodo móvil es un atacante o no. Por lo que el intercambio de direcciones MAC se llevaría acabo entre un nodo legítimo y el atacante.

Sin embargo, aún en esta situación MSP garantiza la privacidad geográfica del nodo legítimo, ya que el atacante no tiene manera de saber si la identidad que recibió en el intercambio le pertenece a dicho nodo o provenía de otro intercambio realizado previamente.

4.4.2. Evaluación de desempeño

Esta sección presenta una comparación de MSP contra estrategias de privacidad similares. Posteriormente, se presentan pruebas para medir el tiempo que añade VIME al procesamiento en cada paquete de datos.

En la literatura, únicamente las referencias [3, 19, 20, 22, 5, 6] operan en entornos LE. Estos trabajos protegen la privacidad geográfica de los nodos móviles ya sea en capa física o en capa MAC. Como puede verse en la tabla 4.6, estas estrategias soportan ataques para la capa que fueron diseñadas; sin embargo, pierden toda utilidad una vez que los atacantes utilizan la información de otra capa. Por ejemplo, en [3], los autores consideran a TPC como una solución potencial al problema de privacidad en capa física. Sin embargo, no es capaz de resistir ataques de MAC *spoofing*. Más aún, como se mostró en el capítulo 3, esta técnica tampoco puede garantizar privacidad geográfica en la capa que fue diseñada. Por otro lado, en esta tabla se muestra que la técnica MSP es capaz de engañar a los atacantes con acceso a la información de ambas capas sin levantar ninguna sospecha.

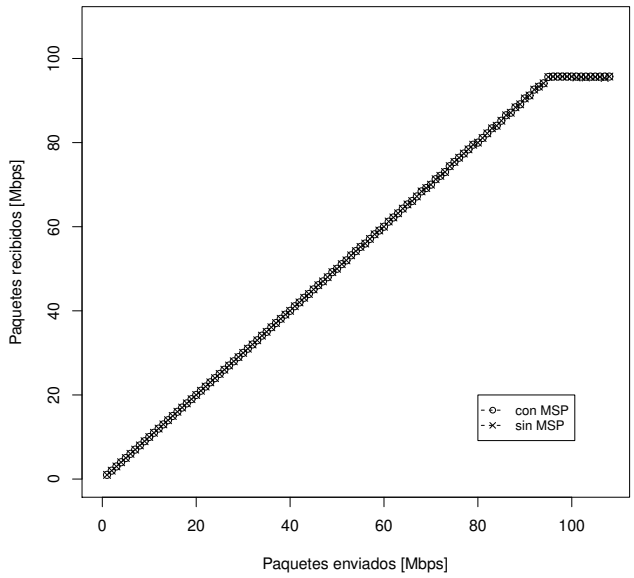
Tabla 4.6: Evaluación de desempeño.

Capa	**Algoritmo**	**Capa física**			**Capa MAC**
		CPS	WCL	MAC *spoofing*	MAC *spoofing*
Física	TPC [3, 19, 20, 22]	✓	×	×	×
MAC	Intercambio de direcciones MAC [5, 6]	×	×	×	✓
Ambas	MSP	✓	✓	✓	✓

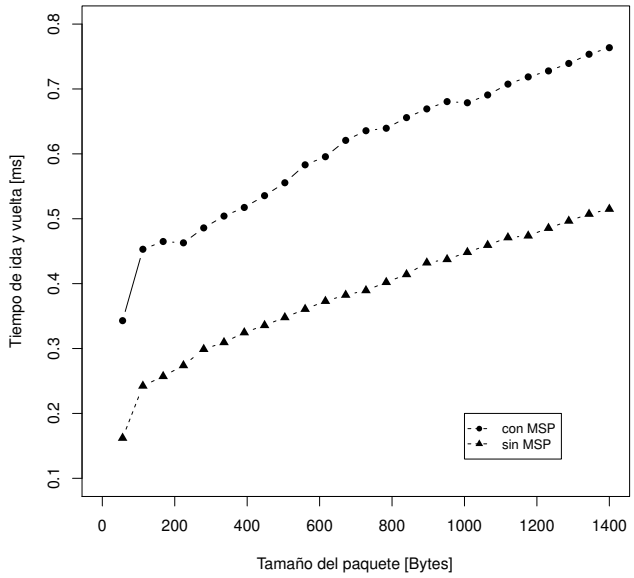
Con el fin de medir el tiempo añadido por VIME al procesamiento de cada paquete de datos, realizamos dos experimentos. En el primer experimento, medimos el desempeño en términos del *throughput*, utilizando la herramienta *Iperf*. Para ello conectamos dos computadoras con sistema operativo Linux que cuentan con una interfaz Ethernet de 100 Mbps. En este experimento, se consideraron tasas de transmisión de datos desde 1 Mbps hasta 108 Mbps. Cada prueba se llevó a cabo 100 veces para obtener valores promedio. Se compararon los resultados de las pruebas con y sin VIME, con el fin de medir la degradación del *throughput*. Los resultados no mostraron ninguna reducción del *throughput* ni pérdida de paquetes. Esto se muestra en la figura 4.11 (a).

En el segundo experimento, estimamos el tiempo que le toma a VIME modificar el encabezado de cada paquete. Para esto utilizamos la herramienta *ping* con el fin de

medir el tiempo de ida y vuelta de los paquetes entre las dos computadoras interconectadas. En esta prueba consideramos tamaños de paquetes entre $56\,\mathrm{B}$ hasta $1400\,\mathrm{B}$. Cada prueba se repitió 100 veces con VIME y 100 veces sin VIME para obtener valores promedio. Los resultados muestran que el tiempo promedio añadido por VIME es de alrededor de $100\,\mu\mathrm{s}$. Esto se puede observar en la figura 4.11 (b).



(a) Mediciones de *throughput*.



(b) Mediciones del tiempo ida y vuelta de ping.

Figura 4.11: Experimentos de desempeño.

4.5. Resumen del capítulo

En este capítulo, primero se modeló el funcionamiento de los atacantes de la capa física a través de un detector que observa mediciones de RSSI en búsqueda de saltos fuera de lo normal que permiten evidenciar un intercambio de identidad. Se propuso una modificación al algoritmo KSD encontrado en la literatura para este fin. Posteriormente se propuso el algoritmo Safe-zone, un mecanismo que le permite a dos nodos móviles intercambiar su direcciones MAC sin dejar rastro para aquellos atacantes que analicen la información de capa física, como KSD. Para lograr esto, el algoritmo Safe-zone le permite a un nodo móvil seleccionar tanto el mejor lugar como el mejor candidato para llevar a cabo un intercambio de direcciones MAC. Por otro lado, con respecto a la capa MAC, se modeló el funcionamiento de los atacantes de la capa MAC a través de dos detectores. El primero de estos detectores considera la información del encabezado IEEE 802.11 con el fin de buscar variaciones poco habituales en el campo número de secuencia con el fin de evidenciar un intercambio de direcciones MAC. El segundo detector considera la información del encabezado IEEE 802.11 con el fin de buscar paquetes de control utilizados en el proceso de asociación de un nodo móvil con un AP. Como contramedida de estos atacantes, en esta tesis proponemos al algoritmo VIME que permite a los usuarios móviles modificar cualquier campo en el encabezado IEEE 802.11, mediante el empleo de interfaces virtuales, antes de entregar los paquetes de datos a la interfaz inalámbrica. VIME garantiza que un atacante que analiza la información de la capa MAC no pueda detectar ningún comportamiento fuera de lo normal, incluso si dos usuarios intercambian sus direcciones MAC simultáneamente. Finalmente, la combinación de los algoritmos Safe-zone y VIME garantiza la privacidad geográfica de los nodos móviles en entornos LE.

Capítulo 5

Conclusiones

En este capítulo, se presentan las conclusiones generales de este trabajo de tesis, la verificación de la hipótesis e interpretación de los resultados tanto del análisis de TPC como del algoritmo propuesto, MSP. Finalmente, se presenta una sección de perspectivas de investigación.

5.1. Conclusiones generales

Contrario a otros trabajos relacionados que consideran a la técnica de ofuscación TPC como una solución potencial al problema de privacidad geográfica en entornos LE, en esta tesis investigamos la viabilidad de usar TPC como una contramedida a los algoritmos de localización empleados por los atacantes. Para lograr este propósito, se desarrolló un modelo probabilístico que permite a un nodo móvil calcular la potencia de transmisión óptima que maximiza la probabilidad de alcanzar solo al AP más cercano, y así incrementar el error de estimación de los algoritmos de localización. Además, en esta tesis se consideraron las condiciones del canal inalámbrico, la densidad de atacantes que escuchan al nodo móvil, así como las limitaciones de los radios Wi-Fi para ajustar la potencia de transmisión óptima. También, se derivaron expresiones analíticas para calcular el error de estimación que generan los algoritmos de localización CPS y WCL. Las simulaciones y pruebas en exteriores mostraron que la eficacia de TPC depende en gran medida de la cantidad de APs que escuchan la señal de los nodos móviles, la capacidad de un nodo móvil para ajustar su potencia de transmisión al valor óptimo y las condiciones del canal inalámbrico. Los resultados confirman que cuanto mayor sea la cantidad de APs, menor será el efecto de TPC en los algoritmos de localización evaluados.

Por otro lado, aun cuando se pueda calcular el valor óptimo de potencia de transmisión (OTPC) que maximiza el error de estimación de los algoritmos de localización, inclusive en escenarios con una alta densidad de atacantes, se mostró que en general los radios Wi-Fi no son capaces de ajustarse a la potencia óptima. Como consecuencia de esto, la protección a los nodos móviles disminuye considerablemente. Más aún, las simulaciones también mostraron que el error de estimación máximo que un nodo móvil puede alcanzar, independientemente del algoritmo de localización empleado por los atacantes, está fuertemente relacionado con las condiciones del canal inalámbrico como son: el exponente de pérdida de energía por la trayectoria y el factor de desvanecimiento. Finalmente, las simulaciones y pruebas en exteriores confirmaron que la eficacia de TPC decrece considerablemente cuando los atacantes implementan algo-

ritmos de localización basados en estrategias *range-free*, como es el caso con WCL. Por ejemplo, para este algoritmo, las simulaciones mostraron que el error de estimación RMS entre OTPC y No-TPC es igual a $19.44\,\mathrm{m}$, mientras que el error de estimación RMS entre LTPC y No-TPC es igual a $2.60\,\mathrm{m}$. Este resultado demuestra claramente la protección limitada que puede alcanzar TPC como una técnica de privacidad geográfica considerando condiciones más realistas.

Asimismo, en esta tesis propusimos al algoritmo MSP, una estrategia que le permite a dos nodos móviles intercambiar sus direcciones MAC evitando la detección por parte de terceros en entornos LE. Esta estrategia es capaz de proveer privacidad geográfica a los nodos móviles aun cuando los atacantes tengan acceso a la información de la capa física y la capa MAC, simultáneamente. Contrario a la mayoría de las propuestas previas, las cuales consideran el intercambio con al menos k nodos móviles, MSP necesita solo dos nodos móviles (i.e. un iniciador y un candidato) para llevar a cabo el intercambio y en consecuencia engañar a los atacantes. Se propuso un algoritmo denominado Safe-zone con el fin de resolver el problema relacionado con la detección del intercambio de direcciones MAC mediante el análisis de la información de la capa física. También, propusimos el algoritmo denominado VIME con el fin de resolver el problema relativo a la detección del intercambio de direcciones MAC. Al combinar ambos algoritmos, se genera una protección integral que evita que los atacantes puedan detectar el intercambio de direcciones MAC ya sea en la capa física o en la capa MAC. Desde el punto de vista de los atacantes, este intercambio disocia la identidad de un nodo móvil de su posición actual. En la sección 4.4 se demostró que cuando el atacante y el algoritmo Safe-zone seleccionan el mismo valor de umbral, Safe-zone tiene una efectividad del $95\,\%$ cuando se lleva a cabo el intercambio de identidad. Más aún, cuando $\mathcal{TH}_{atacante}$ es mayor que $\mathcal{TH}_{safe-zone}$, la eficacia de Safe-zone es casi de un $100\,\%$. Por otro lado, los resultados mostraron que VIME tiene una eficacia del $100\,\%$ con respecto a la detección por parte de terceros en la capa MAC. La combinación de estos dos algoritmos es suficiente para garantizar la privacidad geográfica de nodos móviles. Adicionalmente, nuestros experimentos mostraron que VIME requiere en promedio de $100\,\mu\mathrm{s}$ para procesar cada paquete de datos. Sin embargo, la operación de MSP no mostró ninguna disminución del *throughput*.

5.2. Verificación de la hipótesis

Retomando la hipótesis presentada en la sección 1.2:

"Técnicas de privacidad geográfica basadas en TPC no son efectivas en la práctica debido a factores no estudiados previamente, por el contrario, un solo intercambio de identidad entre dos usuarios que considere la información de la capa física y de la capa MAC tiene la capacidad de proveer de privacidad geográfica a usuarios móviles en un entorno LE".

Suponemos que tanto los AP como los nodos móviles son capaces de medir la potencia de una señal recibida y a su vez estimar la distancia entre su posición y el AP. Además suponemos que el rango de cobertura máximo tanto del AP como de los nodos móviles tiene la misma longitud.

Para verificar la hipótesis previa, esta tesis se desarrolló en dos fases. En la primera, se analizó la eficacia de TPC a través de factores no considerados previamente, tales como: la capacidad de un nodo móvil para ajustar su potencia de transmisión, las condiciones del canal inalámbrico como son: γ, σ y la densidad de AP, y por último

el algoritmo de localización implementado por los atacantes. En la sección 3.4.1 se demostró que la eficacia de TPC está fuertemente relacionada con dichos factores. En específico, en la Figura 3.12 se analizó la interacción de todos estos factores y su efecto sobre el error de estimación de los algoritmos de localización. Más aún, en esta sección mostramos que aun cuando el nodo móvil pudiera satisfacer el requerimiento de potencia de transmisión y mitigar las condiciones cambiantes del canal inalámbrico, se demostró que la eficacia de TPC disminuye considerablemente cuando se usa encontra de algoritmos de localización *range-free*. Por esta razón consideramos que TPC es un algoritmo que no garantiza la privacidad geográfica de los nodos móviles en entornos LE. Esto nos permite concluir que la primera parte de la hipótesis es verdadera.

Por otro lado, en la sección 4.4 se evaluaron tanto los detectores implementados por los atacantes, así como las contramedidas propuestas en esta tesis para proteger la privacidad geográfica de los nodos móviles en entornos LE. La Figura 4.10 muestra el momento cuando dos nodos móviles intercambian sus direcciones MAC utilizando MSP. Esta figura ejemplifica un intercambio de direcciones MAC que no dejó ningún indicio con respecto a los detectores tanto de la capa física como de la capa MAC. Por esta razón consideramos que MSP tiene la capacidad de engañar a los atacantes con acceso a la información de ambas capas de manera simultánea. Por lo anterior, MSP tiene la capacidad de proveer privacidad geográfica a los nodos móviles en entornos LE. Con esto podemos concluir que la segunda parte de nuestra hipótesis también es verdadera.

5.3. Perspectivas de investigación

En el capítulo 4, se mencionó que los atacantes en entornos LE pueden utilizar la información proveniente de capa 3 o superiores tales como: direcciones IP, información de los puertos o estado de los sockets para vincular la identidad del usuario con su posición. Al combinar la información de capa 3 y 4 del modelo de comuniación OSI, los atacantes pueden construir un perfil del usuario (denominado huella), el cual puede ser usado para anular la protección de técnicas de anonimato implementadas en capas inferiores. Para mitigar este problema, se investigará una estrategia que consta de dos pasos. En primer lugar, los nodos móviles seleccionan a un candidato cuya huella es similar. Posteriormente, el nodo móvil debe emular la huella del candidato para pasar desapercibido, y así garantizar su privacidad geográfica en todas las capas del modelo de comunicación OSI.

Como se mencionó en el capítulo 3, derivado del modelo probabilístico que permite calcular la probabilidad TPC_p, se encontró que dicho modelo de probabilidad forma un patrón de diagramas de Voronoi al considerar la topología de los APs de una red inalámbrica. Posteriormente se propondrá un algoritmo que permita a un conjunto de nodos móviles diseñar rutas con la mayor privacidad geográfica posible. Para lograr esto se propone cruzar las aristas de Voronoi de forma perpendicular. Por otro lado, el análisis de probabilidad también puede utilizarse para seleccionar de antemano la técnica de privacidad que esté más acorde a la topología y a las condiciones del canal inalámbrico.

Finalmente, como se mencionó en el capítulo 4, el tiempo que debe durar un seudónimo sigue siendo una pregunta abierta. Aun cuando varias técnicas han sido propuestas en la literatura [32, 55], ninguna de dichas propuestas representa una

solución completa al problema de la frecuencia del intercambio de identidad en redes inalámbricas. En esta tesis optamos por un enfoque más acorde a las necesidades del problema de privacidad geográfica, utilizamos el número de APs escuchando la señal del nodo móvil como métrica para establecer la duración del seudónimo. En específico utilizamos el algoritmo propuesto en [56] para estimar la exactitud con la que los atacantes obtienen la posición de un nodo móvil. De esta manera, entre más atacantes escuchen la señal del nodo móvil, más exacta será su localización, y por consiguiente aumentará la necesidad del usuario de intercambiar su identidad. Como perspectiva de investigación futura se puede investigar más a fondo esta relación y se propondrá una técnica con la cual el nodo móvil pueda determinar de manera oportuna la duración de cada seudónimo de acuerdo a sus necesidades de privacidad geográfica.

Glosario

AoA Angle of arrival.

AP Access points.

GPS Global positioning system.

IEEE Institute of electrical and electronics engineers.

IoT Internet of things.

IP Internet protocol.

LBS Location based services.

LQI Link quality indicator.

MAC Media access control.

OSI Open system interconnection.

PIRE Potencia isotrópica radiada equivalente.

RSSI Received signal strength indicator.

Safe-zone Algoritmo para seleccionar al mejor candidato en un intercambio de direcciones MAC en capa física.

SNG Sequence number gap.

SNRA Sequence number rate analysis.

TCP Transmission control protocol.

TDoA Time diference of arrival.

ToA Time of arrival.

TPC Transmission power control.

VANET Vehicular ad hoc network.

VIME Virtual interfaces MAC address exchange.

Wi-Fi Wireless fidelity.

WLAN Wireless local area network.

Bibliografía

[1] Beresford, A. R. y F. Stajano: *Location Privacy in Pervasive Computing*. IEEE Pervasive computing, 2(1):46–55, 2003.

[2] Krumm, J.: *Inference Attacks on Location Tracks*. Pervasive computing, 4480:127–143, 2007.

[3] Jiang, T., H. J. Wang y Y.C. Hu: *Preserving Location Privacy in Wireless LANs*. En *Procs. of the 5th int. conf. on Mobile systems applications and services*, páginas 246–257, 2007.

[4] Arana, O.: *Técnicas de Privacidad Geográfica en Redes Móviles*. Tesis de Maestría, Universidad Nacional Autónoma de México, Dep. Ingeniería en Telecomunicaciones, Ciudad de México, 2011.

[5] Gruteser, M. y D. Grunwald: *Enhancing Location Privacy in Wireless LAN Through Disposable Interface Identifiers: A Quantitative Analysis*. Mobile Networks and Applications, 10(3):315–325, 2005.

[6] Lei, M., X. Hong y S.V. Vrbsky: *Protecting Location Privacy with Dynamic MAC Address Exchanging in Wireless Networks*. En *IEEE Global Telecommunications Conf. GLOBECOM'07*, páginas 49–53, 2007.

[7] O. Arana, F. Garcia, J. Gomez y V. Rangel: *MSP: Providing Location Privacy in WLAN Networks with a MAC Swapping Protocol*. Computer Networks, 138:136–148, 2018.

[8] Wightman, P. *y cols.*: *Evaluation of Location Obfuscation Techniques for Privacy in Location based Information Systems*. En *IEEE Latin-American Conference on Communications (LATINCOM)*, páginas 1–6, 2011.

[9] Ardagna, C. A. *y cols.*: *An Obfuscation-based Approach for Protecting Location Privacy*. IEEE Trans. on Dependable and Secure Computing, 8(1):13–27, 2011.

[10] Bettini, C. *y cols.*: *Privacy in Location-based Applications: Research Issues and Emerging Trends*, volumen 5599. Springer Science & Business Media, 2009.

[11] Shokri, R. *y cols.*: *Unraveling an Old Cloak: K-anonymity for Location Privacy*. En *Procs. of the 9th annual ACM workshop on Privacy in the electronic society*, páginas 115–118. ACM, 2010.

[12] Lee, K.H. *y cols.*: *ToA based Sensor Localization in Underwater Wireless Sensor Networks*. En *SICE Annual Conference, 2008*, páginas 1357–1361. IEEE, 2008.

[13] J. Xiao, L. Ren y J. Tan: *Research of TDOA based Self-localization Approach in Wireless Sensor Network*. En *International Conference on Intelligent Robots and Systems IEEE/RSJ*, páginas 2035–2040. IEEE, 2006.

[14] Niculescu, D. y B. Nath: *Ad hoc Positioning System (APS) using AOA*. En *Twenty-Second Annual Joint Conference of the IEEE Computer and Communications (INFOCOM)*, volumen 3, páginas 1734–1743. IEEE, 2003.

[15] O. G. Adewumi, K. Djouani y A. M. Kurien: *RSSI based Indoor and Outdoor Distance Estimation for Localization in WSN*. En *IEEE International Conference on Industrial Technology (ICIT)*, páginas 1534–1539. IEEE, 2013.

[16] P. Bahl, V. Padmanabhan y A. Balachandran: *Enhancements to the RADAR User Location and Tracking System*. Microsoft Research, 2(MSR-TR-2000-12):775–784, 2000.

[17] Bauer, K. *y cols.*: *Using Wireless Physical Layer Information to Construct Implicit Identifiers*. 1st Hot Topics in Privacy Enhancing Technologies, 2(2):3–6, 2008.

[18] IEEE Standard 802.11h, http://standards.ieee.org/getieee802/download/802.11h-2003.pdf: *Wireless Medium Access Control (MAC) and Physical Layer (PHY) specifications*, 2003.

[19] Kao, J. y R. Marculescu: *Minimizing Eavesdropping Risk by Transmission Power Control in Multihop Wireless Networks*. IEEE Transactions on Computers, 56(8):1009–1023, 2007.

[20] El-Badry, R. *y cols.*: *Hyberloc: Providing Physical Layer Location Privacy in Hybrid Sensor Networks*. En *Communications (ICC), 2010 IEEE International Conference on*, páginas 1–5. IEEE, 2010.

[21] El-Badry, R. *y cols.*: *Hidden Anchor: A Lightweight Approach for Physical Layer Location Privacy*. En *Journal of Computer Systems, Networks, and Communications*, volumen 2010, páginas 254–258. IEEE Press, 2010.

[22] Taha, S. y X. Shen: *A Physical-layer Location Privacy-preserving Scheme for Mobile Public Hotspots in NEMO-based VANETs*. IEEE Transactions on Intelligent Transportation Systems, 14(4):1665–1680, 2013.

[23] Robitzsch, S., L. Murphy y J. Fitzpatrick: *An Analysis of the Received Signal Strength Accuracy in 802.11 a Networks using Atheros Chipsets: A Solution Towards Self Configuration*. En *IEEE GLOBECOM Workshops (GC Wkshps)*, páginas 1429–1434. IEEE, 2011.

[24] Lu, X. *y cols.*: *Security Estimation Model with Directional Antennas*. En *Military Communications Conference, 2008. MILCOM 2008. IEEE*, páginas 1–6. IEEE, 2008.

[25] Bauer, K. *y cols.*: *The Directional Attack on Wireless Localization-or-how to Spoof your Location with a Tin Can*. En *IEEE Global Telecommunications Conference, GLOBECOM*, páginas 1–6. IEEE, 2009.

[26] Wang, T. y Y. Yang: *Location Privacy Protection from RSS Localization System using Antenna Pattern Synthesis*. En *IEEE Proceedings INFOCOM*, páginas 2408–2416. IEEE, 2011.

[27] Li, X. *y cols.*: *On Modeling Eavesdropping Attacks in Wireless Networks*. Journal of Computational Science, 11:196–204, 2015.

[28] Li, X. *y cols.*: *An Analytical Study on Eavesdropping Attacks in Wireless Nets of Things*. Mobile Information Systems, 2016, 2016.

[29] Shu, T. *y cols.*: *Multi-lateral Privacy-preserving Localization in Pervasive Environments*. En *IEEE Proceedings INFOCOM*, páginas 2319–2327. IEEE, 2014.

[30] Li, M. *y cols.*: *Swing and Swap: User-centric Approaches Towards Maximizing Location Privacy*. En *Procs. of the 5th ACM workshop on Privacy in electronic society*, páginas 19–28. ACM, 2006.

[31] Freudiger, J. *y cols.*: *On Non-cooperative Location Privacy: A Game-theoretic Analysis*. En *16th Procs. of the ACM conf. on Computer and communications security*, páginas 324–337, 2009.

[32] Freudiger, J. *y cols.*: *On the Age of Pseudonyms in Mobile Ad hoc Networks*. En *IEEE Procs. INFOCOM*, páginas 1–9, 2010.

[33] Madory, D.: *New Methods of Spoof Detection in 802.11 b Wireless Networking*. Tesis de Doctorado, Dartmouth College, 2006.

[34] Chen, Y. *y cols.*: *Detecting and Localizing Identity-based Attacks in Wireless and Sensor Networks*. IEEE Trans. on Vehicular Technology, 59(5):2418–2434, 2010.

[35] Li, Q. y W. Trappe: *Light-weight Detection of Spoofing Attacks in Wireless Networks*. En *IEEE Int. Conf. on Mobile Adhoc and Sensor Systems (MASS)*, páginas 845–851, 2006.

[36] Chen, Y. *y cols.*: *Detecting and Localizing Wireless Spoofing Attacks*. En *Securing Emerging Wireless Systems*, páginas 1–18. Springer Science and Business Media, 2009.

[37] *IEEE 802.11 Standard*. http://standards.ieee.org/getieee802/download/802.11-2012.pdf, 2012.

[38] Guo, F. y T. Chiueh: *Sequence Number-based MAC Address Spoof Detection*. En *Int. Workshop on Recent Advances in Intrusion Detection*, páginas 309–329, 2005.

[39] Goldsmith, A.: *Wireless Communications*. Cambridge University Press, 2005.

[40] Garcia, F. *y cols.*: *Ghost: Voronoi-based Tracking in Sparse Wireless Networks using Virtual Nodes*. Telecommunication Systems, 61(2):387–401, 2016.

[41] Vargas-Rosales, C. *y cols.*: *Performance Evaluation of Localization Algorithms for WSNs*. International Journal of Distributed Sensor Networks, 2015, 2015.

[42] Caffery, J.: *A New Approach to the Geometry of TOA Location*. En *52nd Vehicular Technology Conference IEEE-VTS Fall VTC*, volumen 4, páginas 1943–1949. IEEE, 2000.

[43] D. Pompili, T. Melodia y I. F. Akyildiz: *Deployment Analysis in Underwater Acoustic Wireless Sensor Networks*. En *Proceedings of the 1st ACM international workshop on Underwater networks*, páginas 48–55. ACM, 2006.

[44] Al-Alawi, R.: *RSSI based Location Estimation in Wireless Sensors Networks*. En *17th IEEE International Conference on Networks (ICON)*, páginas 118–122. IEEE, 2011.

[45] Mehra, R. y A. Singh: *Real Time RSSI Error Reduction in Distance Estimation using RLS Algorithm*. En *IEEE 3rd International Advance Computing Conference (IACC)*, páginas 661–665. IEEE, 2013.

[46] C. Zang, W. Liang y H. Yu: *The Probabilistic Analysis of Distance Estimators in Wireless Sensor Network*. En *3th International Conference on Natural Computation ICNC*, volumen 5, páginas 270–275. IEEE, 2007.

[47] Zanca, G. *y cols.*: *Experimental Comparison of RSSI-based Localization Algorithms for Indoor Wireless Sensor Networks*. En *Proc. of the workshop on Real-world Wireless Sensor Networks*, páginas 1–5. ACM, 2008.

[48] C. Papamanthou, F. Preparata y R. Tamassia: *Algorithms for Location Estimation based on RSSI Sampling*. En *International Symposium on Algorithms and Experiments for Sensor Systems, Wireless Networks and Distributed Robotics*, páginas 72–86. Springer, 2008.

[49] Sarkar, T.K. *y cols.*: *A Survey of Various Propagation Models for Mobile Communication*. IEEE Antennas and propagation Magazine, 45(3):51–82, 2003.

[50] Yang, J. *y cols.*: *Detection and Localization of Multiple Spoofing Attackers in Wireless Networks*. IEEE Trans. on Parallel and Distributed systems, 24(1):44–58, 2013.

[51] Lubacz, J., W. Mazurczyk y K. Szczypiorski: *Principles and Overview of Network Steganography*. IEEE Communications Magazine, 52(5):225–229, 2014.

[52] Li, Q. y W. Trappe: *Detecting Spoofing and Anomalous Traffic in Wireless Networks via Forge-resistant Relationships*. IEEE Trans. on Information Forensics and Security, 2(4):793–808, 2007.

[53] Krasnyansky, M.: *Universal tun/tap driver*. Online, 2017.

[54] Petit, J. *y cols.*: *Pseudonym Schemes in Vehicular Networks: A Survey*. IEEE communications surveys & tutorials, 17(1):228–255, 2015.

[55] E. Cano Pons, G. Baldini y D. Geneiatakis: *A Wireless Propagation Analysis for the Frequency of the Pseudonym Changes to Support Privacy in VANETs*. En *40th Jubilee International ICT Convention, MIPRO*, páginas 1485–1490, 2017.

[56] Garcia, F. *y cols.*: *LEA: An Algorithm to Estimate the Level of Location Exposure in Infrastructure-Based Wireless Networks*. Mobile Information Systems, 2017, 2017.

[57] Miranda, J. *y cols.*: *Path Loss Exponent Analysis in Wireless Sensor Networks: Experimental Evaluation*. En *11th IEEE Int. Conf. on Industrial Informatics (INDIN)*, páginas 54–58, 2013.

[58] Lin, L., H. C. So y Y. T. Chan: *Accurate and Simple Source Localization using Differential Received Signal Strength*. Digital Signal Processing, 23(3):736–743, 2013.